Article

Balancing the Functionality and Biocompatibility of Materials with a Deep-Learning-Based Inverse Design Framework

Published as part of Environment & Health virtual special issue "Artificial Intelligence and Machine Learning for Environmental Health".

Xiaofang Li, Hanle Chen, Jiachen Yan, Guohong Liu, Chengjun Li, Xiaoxia Zhou, Yan Wang, Yinbao Wu, Bing Yan, and Xiliang Yan*

Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: The rational design of molecules with the desired functionality presents a significant challenge in chemistry. Moreover, it is worth noting that making chemicals safe and sustainable is crucial to bringing them to the market. To address this, we propose a novel deep learning framework developed explicitly for inverse design of molecules with both functionality and biocompatibility. This innovative approach comprises two predictive models and one generative model, facilitating the targeted screening of novel molecules from created virtual chemical space. Our method's versatility is highlighted in the inverse design process, where it successfully generates molecules with specified motifs or composition, discovers synthetically accessible molecules, and jointly targets functional and safe properties beyond the training regime. The utility of this method is demonstrated in its ability to design ionic liquids (ILs) with enhanced antibacterial properties and reduced cytotoxicity, addressing the issue of balancing functionality and biocompatibility in molecular design.

KEYWORDS: *inverse design, biocompatible materials, antibiotic-free strategy, generative models, virtual screening*
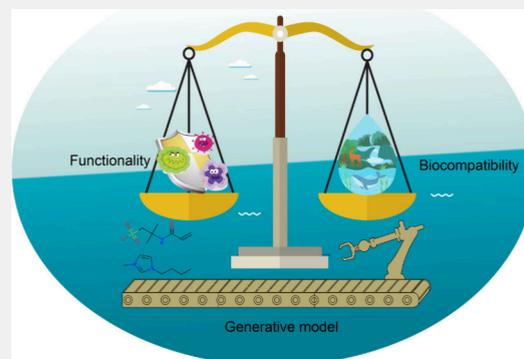
1. INTRODUCTION

Due to their unique physical and chemical properties, ILs have become essential players in many fields of modern science over the past two decades.[1,2] ILs are regarded as "green solvents for the future" because of their low vapor pressure, nonvolatile nature, and good solvation ability.[2−5] The high ionic conductivity, wide electrochemical window as well and fire retardant ability promote ILs as excellent electrolytes in developing more efficient and safer batteries.[6−9] Apart from this, previous studies have shown that ILs are very efficient in various biomedical fields including drug delivery, biosensing, and regenerative medicine.[10−14] On the other hand, since the manufacture and use of ILs are increasing, they are inevitably released into the environment and cause adverse effects on human health.[15] Although the low vapor pressure makes them safe in the atmosphere, ILs are highly persistent in terrestrial and aquatic environments due to their environmental accumulations, excellent miscibility with most media, and high stability.[16,17] In fact, several studies have demonstrated that ILs exhibit toxicity in a broad range of organisms such as proteins,[18] cells,[19] microorganisms,[20] fish,[21] plants,[22] and even mammals.[23] Then, a question arises, "How to design ILs with high functionality but low toxicity"?[24]

Traditionally, trial and error methods generated the desired molecules or materials. However, these methods are time-consuming and labor-intensive. In recent years, artificial intelligence approaches, especially machine learning and deep learning, significantly accelerated the discovery of novel molecules and materials.[25−28] Machine learning for molecular design generally involves two main components, i.e., prediction model construction and virtual screening. For instance, a broad-spectrum antibiotic was identified from over 107 million molecules based on a well-trained deep-learning model.[29] However, there are several limitations of employing the virtual screening: 1) substantial prior knowledge is required to design a virtual compound library, and thus limited candidates can be provided, 2) massive computing resources are needed to explore the vast chemical space, 3) it is difficult to find molecules or materials with better performances from the existing data sets.

To address the above challenges, we proposed a deep-learning-based inverse design framework to generate bespoke molecules with desired properties. Compared with traditional virtual screening, our approach is designed to generate ILs automatically with target properties that avoid brute-force search and computing of every possible IL. Herein, the proposed inverse design framework comprised two main parts, i.e., the generative and predictive models. The generative model could learn a latent representation of the molecule and then generate novel molecular structures. In contrast, the predictive model was used to estimate the properties of every generated molecule. This study applied the proposed framework to design ILs with high antibacterial ability but low cytotoxicity. Finally, a novel IL was identified from passive candidates and experimentally verified. The current study could facilitate the design of environmentally friendly ILs with targeted properties.

2. METHODS

The goal of this study is to design ILs with high antibacterial activity and low toxicity using a deep-learning-based inverse design framework. This framework includes data curation, descriptor calculation, model construction, virtual screening, experimental validation, and molecular dynamics (MD) simulation, as shown in Figure 1. In the inverse design process, the generative model, informed by an extensive library of IL Simplified molecular input line entry specification (SMILES) strings, generates novel and synthetically feasible IL structures. Running parallel, a deep learning predictive

Figure 1. Workflow of our proposed inverse design of biocompatible ILs. The process begins with (1) data extraction for ILs from the ILTox database, followed by (2) the generation of SMILES representations using a generative model with GRUs (Gated Recurrent Units). The generated SMILES structures (3) are input into a predictive model for cytotoxicity and antibacterial activity estimation. At the (4) screening stage, ILs are further evaluated using multiple indicators such as synthesizability and novelty. Subsequently, (5) experimental validation is conducted, and finally, (6) MD simulations are employed to analyze the molecular mechanisms further. IL144 in the figure represents a specific IL instance generated from this workflow with desired properties.

model scrutinizes these structures for specific toxicity end points, such as the minimum inhibitory concentration (MIC) and the half-maximal effective concentration (EC_{50}) values. This dual-pronged approach ensures the selection of ILs that meet predefined toxicity criteria. After the initial virtual screening, selected IL candidates undergo experimental validation to confirm their antibacterial activity and low toxicity levels. Adding to this, MD simulations were used to elucidate the interactions of ILs with bacterial membrane at the molecular level.

2.1. Data Collection and Curation

The GPstack-RNN is a comprehensive model composed of two separate modules: the generative model (G) and the predictive model (P), each trained independently. During the training of the generative model, SMILES strings of 1,183 ILs from the ILTox database (http://www.iltox.com/)[30] were employed as samples. The objective was to train the model to grasp the structural syntax and chemical rules of SMILES.

In the predictive model (P), two data sets were selected from the ILTox database. Data set 1 measures the MIC on the growth of *E. coli* after 24 h and contains toxicity data for 125 ILs. The toxicity values ranged from 0.40 to 5.70 Log10 (μmol/L). Data set 2 is related to the concentration at which ILs achieve 50% of their EC_{50} on HeLa cells over 48 h. It includes toxicity data for 41 samples, ranging from 1.90 to 4.53 Log10 (μmol/L). We employed direct identification of SMILES to construct an end-to-end deep learning model for toxicity prediction of ILs. Toxicity distribution is shown in Figure S1. When using the ILTox database as a sample, the SMILES strings were carefully checked to ensure their correctness. To mitigate the negative impact of feature dimensions on model performance, we employed *z*-score standardization to transform the original MOE descriptors into standard data with a mean of 0 and a standard deviation of 1 (eq 1).

$$z = \frac{x - \overline{x}}{\sigma} \tag{1}$$

Here, for each feature, x represents the raw value, $\overline{x}$ is the population mean, and σ is the standard deviation.

2.2. Construction of the Generative Model

After obtaining SMILES structures of 1,183 chemical molecules from the ILTox database, we pretrained a generative model (G) based on GRU. This process was focused on learning the structural syntax and chemical rules of SMILES, enabling the model to generate chemically viable ILs without any optimization of properties. The GRU network layer consisted of 500 units, with an additional 512 units in the stacked expansion layer and a learning rate set at 0.001. The model underwent training for 1,000,000 epochs, and the changes in the Loss during the training process are depicted in Figure S2. The generative model (G) processes SMILES sequences in two ways: training and generation. During each time step in training mode, the network acquires the training sequence's current prefix and predicts the next character's probability distribution. It then samples the next character from this predicted distribution, compares it with the actual character, calculates the cross-entropy loss, and updates the model parameters. In generation mode, the model parameters are no longer updated. Instead, the network uses the already generated sequence's prefix to predict the next character's probability distribution, which is then sampled and outputted similarly. In this study, the generation count was set to 1,000. The SMILES optimizer in the generative model (G) optimizes the newly generated SMILES structures, using the synthetic accessibility (SA) score in the RDkit library for identification and scoring, filtering out irrational SMILES. Additionally, SMILES that were duplicates of the training set or self-repetitive were excluded, ultimately yielding structurally viable new ILs. We used the t-SNE algorithm to reduce the dimensionality of molecular fingerprint data and visualized them on a two-dimensional plane. These data were generated by RDKit, which provided detailed molecular structure information. This method allows us to intuitively analyze the structural similarities and differences among the ILs. Additionally, we applied the Tanimoto similarity coefficient to quantitatively

analyze the similarity of two ILs. This method measures the similarity between two sets by comparing the ratio of their intersection to their union, with results ranging from 0 (completely different) to 1 (identical).

2.3. Construction of the Prediction Model

The RNN is a deep learning network designed for processing sequential data, capable of using the output from one point in a sequence as the input for the next. In our predictive model (P), we directly employ the SMILES structures of ILs to build an end-to-end model. First, SMILES are standardized to ensure the uniqueness of molecular representations, aiding in the more accurate identification of ILs' molecular structures. Then, characters within the SMILES are transformed into ASCII numeric matrices (Figure S3) and processed through GRU network units, aligning them with corresponding toxicity data, as depicted in Figure 1. Data sets about *E. coli* and HeLa cells were selected to train an efficient predictive model. The model parameters are set as follows: a hidden_size of 200, 2 num_layers, 500 epochs, a batch size of 16, and a learning rate of 0.001. The loss function is the root-mean-square error (RMSE), and the optimizer is the Adam algorithm. Additionally, MOE descriptors and two machine learning methods (Random forest (RF) and XGBoost) were utilized for QSAR modeling of the data sets, with a model assessment conducted through 5-fold cross-validation. Since our model is intended for external validation through virtual screening, we only conducted 5-fold cross-validation. The evaluation metrics for regression models included coefficient of determination (R^2), RMSE, and mean absolute error (MAE), which were calculated according to the following eqs 2, 3, and 4

$$R^2 = \frac{\sum_{i=1}^n (y_i^{obs} - y_i^{pred})^2}{\sum_{i=1}^n (y_i^{obs} - \bar{y}_i^{obs})^2} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i^{obs} - y_i^{pred})^2}{n}} \quad (3)$$

$$MAE = \frac{\sum_{i=1}^n |y_i^{obs} - y_i^{pred}|}{n} \quad (4)$$

where y_i^{pred} is the predicted value for each specific molecule, y_i^{obs} is the experimental value for each specific molecule, $\bar{y}_i^{obs}$ is the average value across all molecules, and n is the number of molecules.

The range of hyperparameters adjusted in machine learning and deep learning is shown in Table S1. Additionally, we utilized the SHapley Additive exPlanations (SHAP) algorithm to analyze the feature importance of the RF model and visualized the RNN model through the class activation feature method.[31] SHAP assigns an importance value to each feature to explain its contribution to the model's prediction outcome. Class activation features are visualized by displaying detection results of all channels in different colors, where the intensity of the color reflects their contribution. Moreover, ILs with effective antibacterial properties and low cell impact were selected from the *E. coli* and HeLa cell data sets, with the prediction accuracy validated through IL synthesis and external experiments.

2.4. Synthesis of the Novel IL Screened from Inverse Design

To design environmentally friendly ILs, this study utilized the GPstack-RNN deep learning network to select new ILs with high antibacterial potential and low cytotoxicity, exemplified by IL144, which was experimentally validated. *E. coli* strain (ATCC25922) was obtained from the Beijing Municipal Culture Collection (BJMCC), and the HeLa cells were purchased from the ATCC. The compound IL144 was synthesized experimentally by the Lanzhou Institute of Chemical Physics, China. Specifically, IL144 was synthesized from trioctylphosphine (molecular weight 370.64, density 0.831 g/mL), iodomethane (molecular weight 141.939, density 2.28 g/cm³), and potassium bis(trifluoromethylsulfonyl)imide (molecular weight 319.244). The synthetic reactions were carried out according to eqs 5 and 6.

$$P(C_8H_{17})_3 + CH_3I \rightarrow CH_3P^+(C_8H_{17})_3I^- \quad (5)$$

$$CH_3P^+(C_8H_{17})_3I^- + KTFSI \rightarrow CH_3P^+(C_8H_{17})_3TFSI^- + KI \quad (6)$$
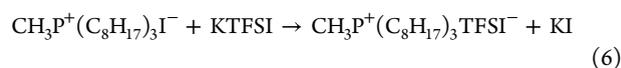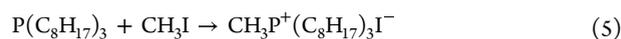
The synthesis process is as follows: 1) Synthesis of the intermediate methyltrioctylphosphonium iodide: Under nitrogen protection, 8.92 mL (0.02 mol, 7.41 g) of trioctylphosphine was mixed with 100 mL of anhydrous toluene. The mixture was stirred vigorously in an ice bath, and 1.25 mL (0.02 mol, 2.84 g) of iodomethane was added dropwise. The reaction was stirred in the ice bath for about 3 h, then heated to 60 °C and stirred for another 1.5−2 h. After completion, the reaction mixture was cooled to −20 °C overnight, then subjected to vacuum filtered and washed with anhydrous ether to yield a pale yellow solid crude product. The intermediate was purified twice by recrystallization from acetone-ethyl acetate and washed with anhydrous ether, yielding a light yellow methyltrioctylphosphonium iodide. This intermediate was then transferred to a sealed container and stored under nitrogen. 2) Synthesis of the IL: Weigh 5.13 g (0.01 mol) of the methyltrioctylphosphonium iodide intermediate and 3.19 g (0.01 mol) of potassium bis(trifluoromethylsulfonyl)imide, dissolve in 100 mL of high-purity water, and place in a separatory funnel for vigorous shaking. After settling, wash three times with high-purity water to remove the potassium iodide byproduct. The water in the IL was removed using a rotary evaporator, followed by vacuum heating for drying, then transferred to a sealed container and stored under nitrogen. The resulting methyltrioctylphosphonium bis-(trifluoromethylsulfonyl)imide salt was a light yellow viscous liquid with a yield of 45%.

2.5. Antibacterial Assay

The antibacterial activity of *E. coli* was determined using the tube dilution method. First, *E. coli* was cultured in Mueller-Hinton broth at 37 °C for 24 h. Subsequently, bacterial suspensions with a concentration of 106 cfu/mL (cfu, colony-forming units) were prepared from each culture. Solutions of ILs (31.3 to 1500 mmol/L) were prepared and sterilized further through a 0.45 mm membrane filter. Then, in a 96-well microplate, varying concentrations of bacterial suspensions (50 μL) were mixed with IL solutions (50 mL). The optical density at 600 nm (OD600) was measured using a fluorescence spectrophotometer, with vigorous shaking to ensure proper incubation of bacteria, and readings were taken every 10 min for a total period of 24 h to establish bacterial growth curves (Figure S4). Control experiments were also conducted, including a negative control without microorganisms and a positive control without ILs.

2.6. Cytotoxicity Assay

To assess the cytotoxicity of ILs on HeLa cells, the following methodology was employed: HeLa cells (ATCC official Web site) were cultured in 25 cm² Corning flasks using high-glucose Dulbecco's Modified Eagle Medium (DMEM, Hyclone) supplemented with 10% fetal bovine serum (FBS, Hyclone), under a controlled environment of 37 °C and 5% CO₂ in a humidified incubator. A suspension containing 5000 HeLa cells in 100 μL was seeded into each well of a 96-well plate and incubated for 24 h. Subsequently, the cells were exposed to a range of IL concentrations (90.8, 181.6, 227, 302.7, 363.2, 454, 605.3, 908, and 1816 mg/L) for 48 h, with five replicate wells per concentration. The plate included blank controls, untreated controls, and a series of diluents, each replicated at least three times and a total volume of 100 μL per well. After incubation, the plate was left to stand at room temperature for 10 min and then washed twice with 200 μL of PBS. In the darkness, 50 μL of serum-free medium and CellTiter-Glo reagent were added to each well. The samples were then placed on a shaker (150 r/min for 2 min) to promote cell lysis and allowed to stand for 10 min. Luminescence intensity was measured at a wavelength of 590 nm using a fluorescence enzyme marker (BIOTEK Synergy HTX). Data fitting was performed using Micromath Scientist 2.0 software to plot growth curves and subsequently calculate the EC₅₀ (Figure S5).
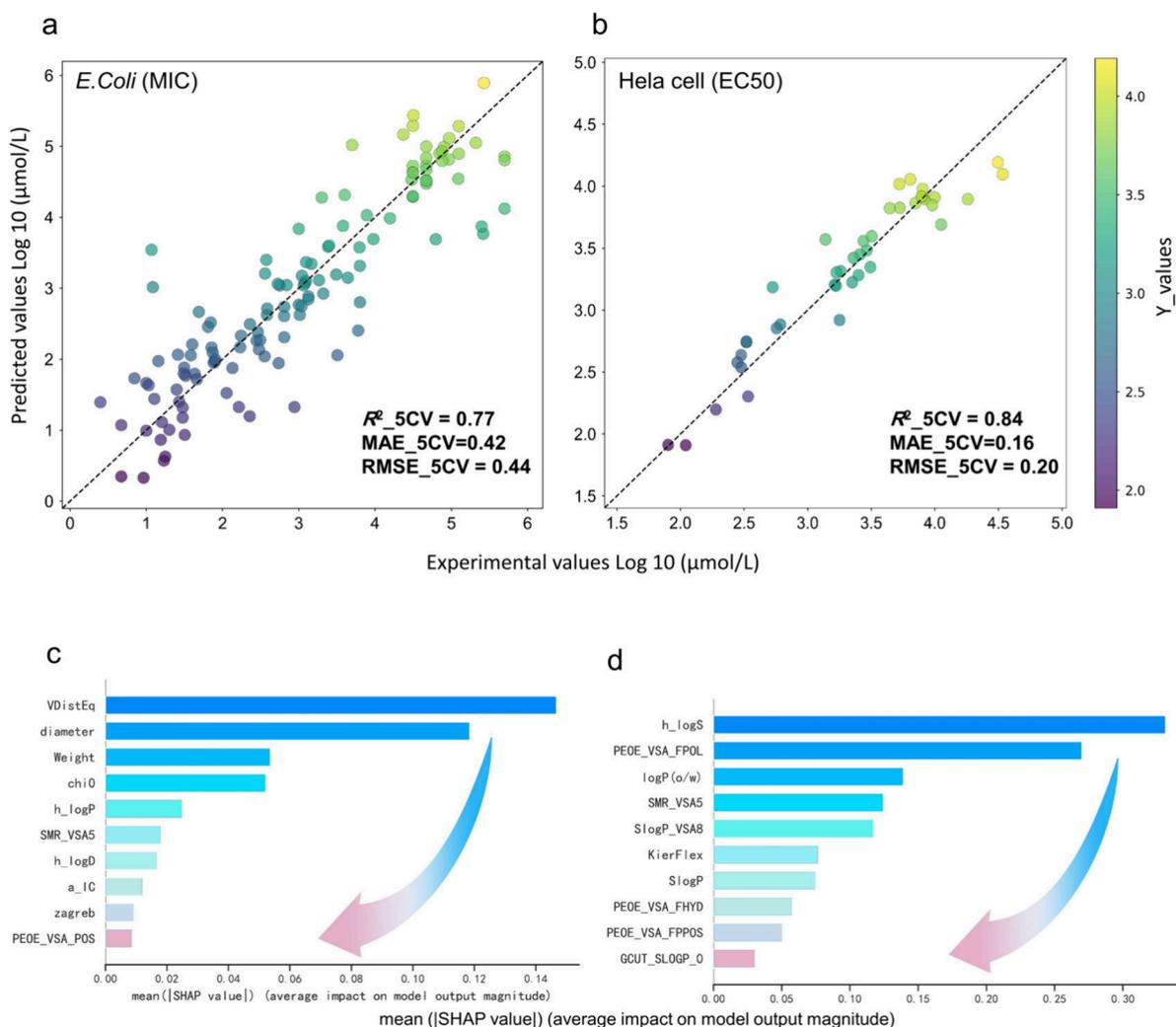
Figure 2. Predictive performance and feature importance analysis of the machine learning models on two data sets. Correlations between the predictions from RNN models and the experimentations of (a) the *E. coli* (MIC) data set and (b) the Hela cell (EC$_{50}$) data set. Ranking of important features calculated from SHAP value analysis of RF models on (c) the *E. coli* (MIC) data set and (d) the Hela cell (EC$_{50}$) data set. R^2, RMSE, and MAE were also shown in the figure.

2.7. MD Simulations

The interactions of ILs with cell membranes were simulated using MD methods. To further investigate the antibacterial mechanism of ILs, we selected IL144 for mechanistic studies. The topology and force field parameters of IL144 were generated using CHARMM-GUI, based on the CHARMM General Force Field (CGenFF), detailed information can be found in the Method S1.[32] The model *E. coli* membrane consisted of three main components: POPE (72%), POPG (23%), and POCL1 (5%), following compositions widely used in prior simulations.[33,34] Water was modeled with the 3-site rigid TIP3P model and sodium ions were added for charge neutralization. The initial size of the simulation box was 11 nm × 11 nm × 10 nm. The IL144 cation and anion were positioned above the equilibrated bilayer to set up the system.

Energy minimization using the steepest descent method was performed until the maximum force converged below 1000 kJ/mol. Then, the system underwent equilibration in NVT (constant Number of atoms, Volume, and Temperature) and NPT (constant Number of atoms, Pressure, and Temperature) ensembles, detailed information can be found in the Method S1. After that, a 200 ns production MD simulation was performed based on the equilibrated system in the NPT ensemble. The temperature was fixed at 310 K using the Nose-Hoove method with a coupling constant of 1 ps. The pressure was kept at 1 bar using a semi-isotropic Parrinello−Rahman barostat with

a coupling constant of 1 ps. The Lennard-Jones parameters for nonbonded interactions were determined using the conventional Lorentz-Bertelot combination rules. All nonbonded interactions were truncated at a cutoff of 1.2 nm, and the long-range electrostatic interactions were calculated using the particle-mesh-Ewald method. The covalent bonds were constrained using the Lincs algorithm. All simulations were performed with a time step of 2 fs using GROMACS v2022.2.[35] Periodic boundary conditions were considered in all three directions of the simulation system. The simulation snapshots were drawn using visual molecular dynamics (VMD).[36]

3. RESULTS AND DISCUSSION

3.1. The End-to-End Deep Learning Models Exhibited Excellent Performances for Predicting Antibacterial Activity and Cytotoxicity of ILs

This study directly utilized SMILES strings to construct QSAR predictive models, replacing traditional descriptors. To further demonstrate the advantages of SMILES-based identification of ILs, MOE descriptors and two machine learning methods (RF and XGBoost) were respectively used to develop QSAR models. As shown in Figure 2 and Table 1, the end-to-end deep learning models exhibited superior performance to

Table 1. Predictive Performance of QSAR Models Constructed by Traditional Machine Learning (MOE Descriptors) and End-to-End Deep Learning (SMILES) Algorithms

Database	MOE		SMILES
	RF	XGBoost	RNN
E. coli	0.63 (0.58)[a]	0.64 (0.62)	0.77 (0.44)
Hela	0.71 (0.36)	0.68 (0.40)	0.84 (0.20)

[a]Values in the brackets represent the RMSE of each model.

traditional machine learning models. For the *E. coli* data set, the RNN deep learning model achieved a predictive R^2 of 0.77 (RMSE = 0.44), versus 0.64 (RMSE = 0.62) for the XGBoost model. Similarly, the RNN model attained a predictive R^2 of 0.84 (RMSE = 0.20) on the Hela data set, compared to 0.70 (RMSE = 0.36) for the optimal machine learning model. The comparable R^2 and RMSE values from cross-validation and external validation across all models indicated that the constructed machine learning models performed consistently

on different data sets, demonstrating good robustness and generalization capability. All coefficient of determination (R^2) values exceeded 0.61, meaning the machine learning models successfully captured the relationships between IL structures and associated toxicity end points. Compared to traditional machine learning modeling using molecular descriptors, directly encoding IL SMILES strings as input to construct QSAR predictive models via RNN demonstrated superior predictive performance in this study. This can be attributed to several factors: 1) SMILES strings contain rich chemical information, enabling effective learning of molecular features through end-to-end training; 2) sequence inputs retain molecular graph topology, better capturing inherent structure−activity relationships; 3) deep learning models have stronger capabilities for fitting complex mappings. This sequence-based modeling framework has advantages like stronger generalization, laying the foundation for developing universal platforms to predict compound bioactivities.

Feature importance analysis of the predictive models allowed us to identify several structure features or phys-
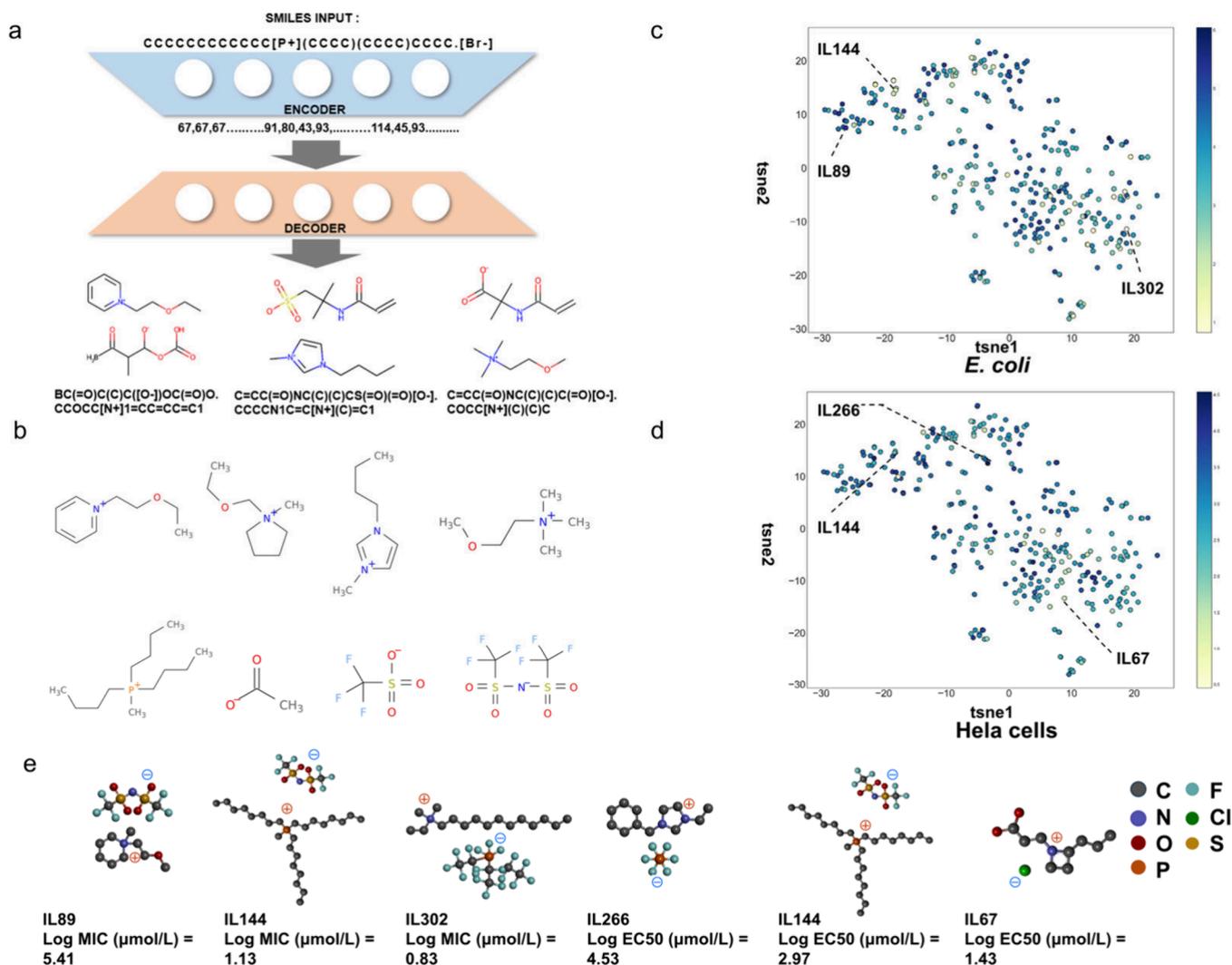
Figure 3. Generative model created diverse novel ILs. (a) The SMILES-based ILs are transformed into ASCII through an encoder, and the structure of a new IL is generated from the decoder. (b) Molecular structure diagram of representative ILs. (c, d) The t-SNE distribution of 342 newly generated ILs produced. Each dot represents an IL, and the color indicates the antibacterial activity and cytotoxicity predicted from the deep learning models. (e) Six representative ILs were selected from the newly generated chemical space. They were predicted to have high, medium, or low antibacterial and toxic properties, respectively.
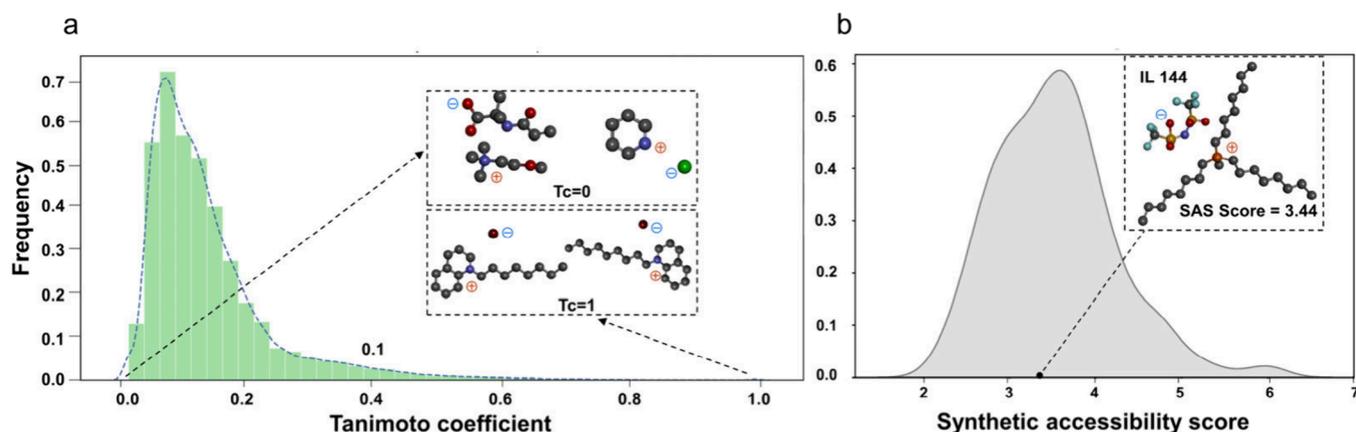
Figure 4. Novelty and synthesizability evaluation for newly generated ILs. (a) The Tanimoto similarity between the newly generated ILs and those collected from the ILTox database. The higher the similarity, the greater the Tc score. Most generated ILs are dissimilar from the molecules in the ILTox database. The figure shows two extremely different (IL5 and IL30) and essentially identical (IL4 and IL118) ILs. (b) The SA score distribution of the generated new IL structure; most ILs are easily synthesizable. IL144 showed high antibacterial activity (*E. coli*) and low cytotoxicity (Hela cells).

icochemical properties responsible for antibacterial activity and cytotoxicity, which can be used to elucidate potential mechanisms for these bioactivities and thus guide the design of novel ILs. The ranking of molecular descriptors was calculated from SHAP value analysis of the optimal RF models, as shown in Figures 2c and d. The high ranking of a molecular descriptor indicates its pivotal role in the final predictive model. In Figure 2c, the diversity of atomic distances within a molecule (VDistEq), molecular diameter, molecular weight, the complexity of its structural connections (chi0), and its propensity to distribute between oil and water environments (h_logP) have been identified as key factors that influence the prediction of a compound's antibacterial activity. VDistEq, by analyzing how atoms are spaced within a molecule, significantly impacts the compound's effectiveness in binding to and acting against bacteria. The molecular diameter is essential in determining the compound's capability to penetrate bacterial cell membranes, influencing how effectively it can reach its target. Collectively, VDisteq and molecular diameter are crucial in determining a compound's antibacterial properties. In Figure 2d, several key parameters are emphasized for their importance in predicting cytotoxicity. H_logS indicates the solubility of a compound in water, PEOE_VSA_FPOL shows the size of the charged (polar) portion of a molecule, and logP (o/w) measures the compound's preference for distribution between oily and aqueous environments. Additionally, SMR_VSA5 and SlogP_VSA8 reflect the size of specific regions on the molecular surface and the area of lipophilic regions, respectively. Among these, h_logS plays the most crucial role in the mechanism of antibacterial action. A higher h_logS value typically means that the compound is more soluble in water, which aids in its more effective penetration through bacterial cell membranes to reach its site of antibacterial action. On the other hand, PEOE_VSA_FPOL is also vital for the binding of the molecule to bacterial cells. Its size and distribution directly affect the antibacterial efficacy of the molecule, as this determines how effectively it can interact with key sites on the cell. The analysis of feature importance is instructive for understanding and optimizing the interaction of ILs with biological systems, guiding the enhancement of safety and efficacy.

Additionally, we demonstrated how stack-RNN learns and memorizes information while processing SMILES strings. We analyzed the activations of neurons within the neural network during data processing. Figure S6 shows a sample IL with interpretable activations in hidden layers, with each line representing neuron activations at different processing steps, where colors change from dark blue to dark red indicating activation values from −1 to 1. As depicted in the color map, our RNN model contains two types of interpretable neurons: chemically sensible neurons and syntactically sensible neurons. The chemically sensible neurons activate in the presence of specific chemical moieties, and the syntactically sensible neurons monitor the syntactic groups of SMILES such as numbers and brackets. Although neural-network-based models are typically hard to interpret, the partial interpretability provided by this method can help understand how the models make prediction and is highly valuable to researchers.

3.2. Novel IL Structures Were Created from stack-RNN Generative Models

In this study, we utilized a stack-RNN to construct a generative model (G) capable of producing 342 novel ILs with diverse chemical properties, demonstrated by the detailed SMILES representations available in Table S2. This model, comprising an encoder-decoder architecture, efficiently transforms input SMILES sequences of ILs into molecular latent vectors, which are then decoded back into new SMILES sequences (Figure 3a). Notably, the generated ILs, including various types such as imidazoles and pyridines, showed a broad diversity, indicating a substantial expansion of the chemical space (Figure 3b).

342 novel ILs contain 233 cations and 86 anions. To visually demonstrate the structural features of the ILs, we analyzed several representative compounds. Among these cations, quaternary ammonium compounds dominate due to their diversity, including simple quaternary ammonium cations, cyclic quaternary ammonium cations with closed-ring features, and quaternary ammonium cations containing heterocyclic elements. Moreover, our generative model has also produced a variety of other types, such as ammonium cations and their derivatives, pyrrolidinium cations, and imidazolium cations. Furthermore, the cation structures generated by the model include simple inorganic cations, such as Na^+. These categories
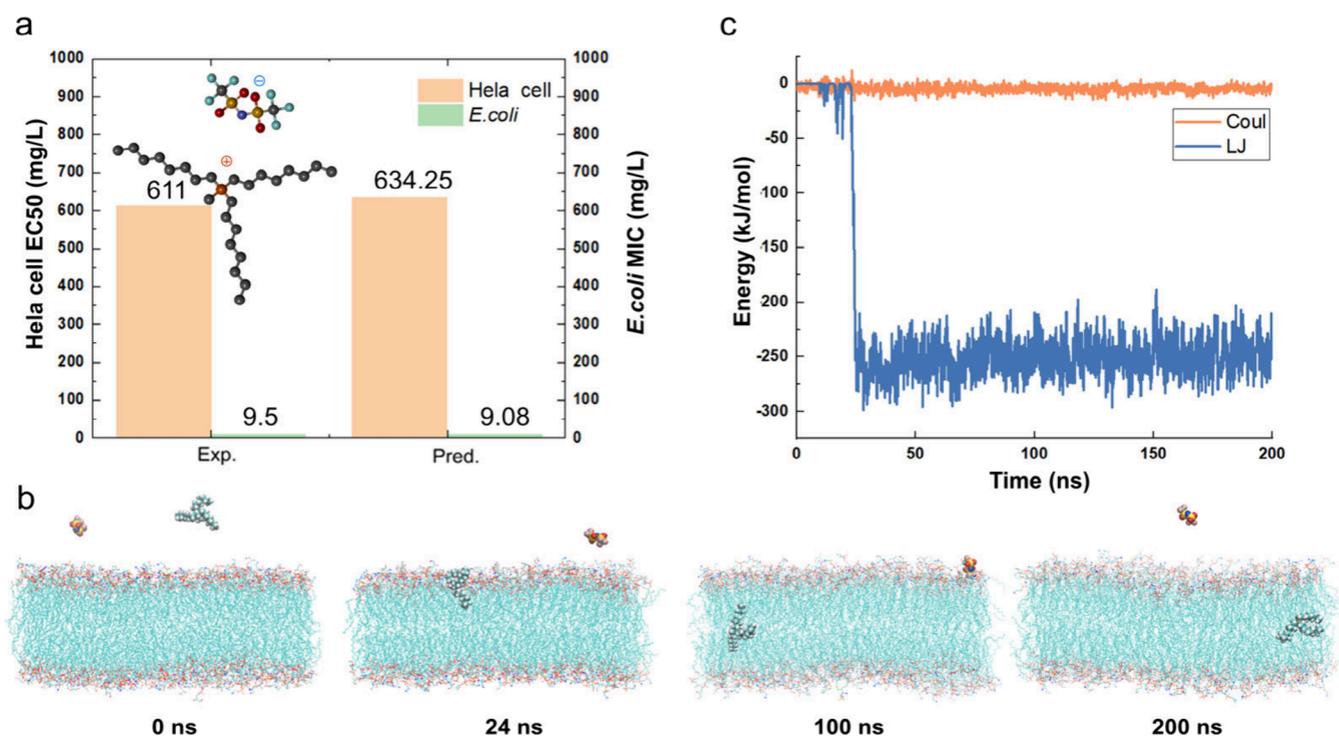
Figure 5. A novel IL was identified from our deep learning framework. (a) Comparison of predicted and experimental values of IL144. The selected IL144 contained a $P(C_8H_{17})_3CH_3$ cation and a $N(SO_2CF_3)_2$ anion. (b) MD simulations to clarify the binding modes of IL 44 with the cell membrane. (c) The average energies of both van der Waals (LJ) and electrostatic (Coul) interactions between the IL144 cation and the membrane.

of ILs have distinct advantages in chemical properties and applications. For example, quaternary-ammonium-based ILs, recognized for their extensive antibacterial properties, have been extensively studied in various fields, including disinfectants, surfactants, and herbicides. Imidazolium-based ILs are currently being developed into new antimicrobial drugs due to their potential antibacterial, antifungal, and anticancer properties and ecological toxicity effects. Additionally, pyridinium-based ILs are also under investigation as potential anticancer agents.[37−41] In the array of anions, those containing sulfates and halides are the most abundant, followed by acetate and other complex fluoride-containing anions. The diversity of these anions expands the potential for the development and application of ILs across various fields. For instance, ILs comprising sulfate and halide anions exhibit efficacy in antimicrobial and anticancer applications.[42−45]

To evaluate the structural diversity of the generated ILs, we calculated their distribution in chemical space. We used t-SNE dimensionality reduction to map 342 ILs onto a two-dimensional plane. The chemical space was filled with predicted toxicity data for *E. coli* (Figure 3c) and Hela cells (Figure 3d) as biological end points, with different colors marking the values. Six ILs, each with high, medium, or low antibacterial and toxic properties were presented in Figure 3e. As shown in Figure 3, the generated ILs were widely distributed across the chemical space, covering various types of ILs containing imidazolium, pyrrolidinium, quaternary ammonium, and phosphonium cations. This indicates that the stack-RNN model could effectively learn the latent representations of molecules and generate novel ILs with rich structural diversity. Additionally, ILs with similar toxicity tended to cluster together in the chemical space, consistent with the rules of structure−activity relationships. ILs with

analogous chemical structures and physicochemical properties also exhibited approaching bioactivity.[46−48]

In addition, to quantify the novelty of the generated ILs, we employed the Tanimoto coefficient (Tc) to compare chemical similarities between the virtual library from our stack-RNN model and existing compounds in the ILTox database. As shown in Figure 4a, a total of 583,11 distance values ranging from 0 to 1 with an average of 0.13 were produced among the 342 ILs. Generally, two molecules with a Tc greater than 0.5 are considered structurally similar. The lower the Tc value, the more dissimilar the structures of the two molecules.[49] For example, IL71 and IL264 had different cation and anion types, with a Tc of 0, and thus were deemed structurally dissimilar. The Tc distribution was mainly concentrated between 0 and 0.4, accounting for 96.2%, indicating that most generated ILs were structurally distinct from those in the training set and exhibited high novelty. Therefore, the ILs generated by GPstack-RNN possessed high structural diversity, which increased the probability of discovering novel ILs with high antibacterial activity and low toxicity.

In addition to structural evaluation, synthetic feasibility is an important criterion for assessing the "quality" of inverse designed molecules. Therefore, we utilized the SA score method to evaluate the plausibility of the newly designed IL structures. The SA score is calculated by weighting the frequency of ECFP4 fingerprints obtained from 10 million compounds in PubChem, allowing rapid assessment of synthetic difficulty for many compounds.[50,51] As shown in Figure 4b, the SA scores of the ILs in this study ranged from 2.11 to 6.09 with an average of 3.53. Higher SA scores indicate greater difficulty in synthesis, and molecules with SA scores above 6 are generally considered infeasible to synthesize.[52] In this study, 99.1% of the new molecules had SA scores below 6, with only 3 molecules slightly higher than 6 but close to it.

Therefore, although the newly generated ILs exhibited novelty, most designed compounds were deemed synthetically accessible.

In summary, this study evaluates the usefulness and success of a generative model optimized with a stack-RNN (G) based on several criteria: (1) molecular diversity and novelty evaluated by the chemical space distribution and structural similarity analysis; (2) chemical synthesis ability measured with the SA score; (3) learning progress tracked by cross-entropy loss during model training; (4) model performance confirmed by virtual screening and experimentation. These evaluations demonstrate that the generative model has successfully created novel, synthetically feasible candidate compounds, providing the potential for further discovery of ILs with enhanced functionalities.

3.3. A Desired IL with High Antibacterial Activity and Low Cytotoxicity Was Identified and Experimentally Validated

The screening objective of this study was to obtain ILs with antibacterial activity and low cytotoxicity. Therefore, two main screening criteria were established: first, the ILs should inhibit the growth of typical bacteria; second, the ILs should have less toxic effects on human hosts. In addition, the feasibility of IL synthesis was also considered, and those molecules with lower synthesis difficulty were preferred. For Hela cells, the EC_{50} value of IL144 was predicted to be 634.25 mg/L, while the predicted MIC value for *E. coli* was 9.08 mg/L. Additionally, the SA score of IL144 was calculated as 3.44, indicating that the compound has high antibacterial activity and low toxicity and is easy to synthesize. According to these screening criteria, a novel compound, IL144, was finally selected from the constructed IL library.

In the structure of IL144, phosphorus serves as the central atom of the cation. Specifically, phosphorus-containing ILs, due to their lower hydrophobicity (i.e., higher polarity), reduce bacterial tolerance to quaternary phosphate salts, thus exhibiting superior antibacterial properties structurally.[53] In the context of anions, while fluorine-containing anions might contribute to cytotoxicity through the release of fluoride ions, their overall impact on the toxicity of ILs is comparatively minor.[54,55] Therefore, given the structural features of IL144, it is expected to demonstrate high antibacterial activity and low cytotoxicity. The results of the nuclear magnetic resonance (1H NMR) (Figure S7) and electrospray ionization mass spectrometry (ESI-MS) for IL144 are as follows: 1H NMR (400 MHz, DMSO-d6) δ (ppm) = 2.15 (t, 6H, P-CH2-), 1.90 (t, 3H, P-CH3), 1.51−1.23 (m, 36H, -CH2-), 0.91 (t, 9H, -CH3); $C_{27}H_{54}F_6NO_4PS_2$; ESI-MS: m/z 666.3 (m+1)+.

As shown in Figure 5a, we compared the predicted and experimental values of the ILs screened by the GPstack-RNN model. Overall, there were minor differences between the predicted and experimental values, further indicating the robustness and generalizability of our constructed machine learning models. The predicted EC_{50} value for Hela cells was 634.25 mg/L, while the experimental value was 611 mg/L; the predicted MIC value for *E. coli* was 9.08 mg/L, while the experimental value was 9.5 mg/L. The experimental results verify that IL144 has a good antibacterial effect and low cytotoxicity. This process demonstrates the effectiveness of computationally driven targeted screening strategies. Furthermore, a good model depends on the high quality of data rather than just the size of the data set. Although the data set used in this study is not large (125 and 41 data points in two data sets), the model still demonstrated excellent predictive capability through experimental validation.

In this study, IL144 exhibited different effects on *E. coli* and HeLa cells. For *E. coli*, IL144 directly disrupts the bacterial cell wall or membrane, causing rapid cell death, and thus shows significant antibacterial activity at lower concentrations.[56,57] In contrast, the complex eukaryotic structure of HeLa cells requires IL144 to act through more intricate intracellular pathways, such as affecting cell signaling, inducing apoptosis, or interfering with the cell cycle, necessitating higher concentrations for similar effects.[58,59] The differences in EC_{50} values primarily stem from the structural and permeability differences between the two cell types.

3.4. MD Simulations Revealed the Antibacterial Mechanisms of the Newly Discovered IL

In the MD simulations, Figure 5b illustrates the dynamic interaction between the IL144 cation−anion pair and the cell membrane. Initially, at 0 ns, the IL144 cation is positioned externally to the membrane. The simulation reveals that, progressing toward 24 ns, the cation gradually approaches and commences interaction with the membrane's surface. By the 100 ns mark, we observe partial insertion of the IL144 cation into the membrane. This insertion progresses, between 100 and 200 ns, the cation fully integrates into the membrane, signifying a stable binding interaction. The insertion of the IL144 cation into the cell membrane at 100 ns can be attributed to its relatively high lipophilicity, facilitating interaction with the hydrophobic tails of the phospholipid molecules in the membrane. This characteristic of lipophilic cations is known to disrupt the structural integrity of bacterial membranes, potentially leading to bacterial cell death or growth inhibition. Furthermore, the cation's integration into the membrane could modify the membrane potential, influencing ion channels and transport proteins. This alteration may have consequential effects on the normal physiological functions of the cell. In conclusion, the IL144 cation's distribution and interaction within the membrane are crucial for understanding its toxicity mechanism and are directly related to its antibacterial potency.

Concurrently, during the 0 to 200 ns simulation period, the IL144 anion consistently remained external to the cell membrane. This observation can likely be attributed to its higher hydrophilic nature and the electrostatic repulsion it experiences from the negatively charged phospholipids of the membrane. Anions typically have difficulty penetrating cell membranes which are composed of lipophilic molecules. Therefore, the IL144 anion's biological activity could occur through interactions with positively charged proteins or other molecules on the membrane surface rather than through membrane penetration.

Figure 5c presents the time-based progression of the average energies associated with van der Waals and electrostatic interactions between the IL144 cation and the cell membrane. During the early stages of the simulation, the dramatic fluctuations in van der Waals energy correspond to the dynamic movements of the IL144 cation as it progressively nears the cell membrane. Over time, these fluctuations in van der Waals energy diminish, indicating a progression toward a more stable interaction between the cation and the membrane. In contrast, the electrostatic energy remains relatively stable throughout the simulation, indicating a consistent charge interaction between the IL144 cation and the membrane. This

suggests that electrostatic interactions are pivotal in the binding process. This stable electrostatic interaction is a key factor in determining whether the IL144 cation can effectively penetrate and establish a stable binding with the membrane. Figure S8 illustrates the energy changes between IL144 anions and the cell membrane over time. Coulomb energy remains stable and slightly negative, indicating balanced charge interactions. Van der Waals energy shows major fluctuations, particularly between 100 to 200 ns, dropping sharply below −800, suggesting strong repulsion due to changes in distance between the anion and the membrane.

We can conclude that this interaction is a complex process through MD simulation and energy analysis of the interaction between IL144 and the cell membrane. It encompasses initial contact, exploration of binding sites, and, ultimately, penetration or stable binding. The energy analysis offers a dynamic view of this process, revealing the variations in interaction forces over time and molecular position. These insights deepen our understanding of the antibacterial mechanism of IL144 and lay a foundational scientific groundwork for developing more efficacious ILs.

4. CONCLUSIONS

In this study, we proposed a novel strategy for the inverse design of environmentally friendly ILs, termed GPstack-RNN. This method integrates a generative model (G) and a predictive model (P) based on deep learning, which, after separate training, are jointly used to generate ILs with targeted characteristics (high antibacterial and low cytotoxicity). Using GPstack-RNN, we screened several novel ILs with high antibacterial and low toxicity, among which IL144 was experimentally validated as an example. Experimental results show that the generative model can produce ILs with the desired properties beyond the training data and possess strong generalization capabilities. By learning the structure and chemical rules from existing data, the generative model creates diverse molecular structures, expanding the chemical space and enhancing the likelihood of discovering new functional materials. The same basic approach can be applied to other types of functional material design, such as anticancer drugs, which could increase the discovery rate and decrease associated costs. Finally, perhaps the most useful benefit of using our proposed framework to create molecules is that they can offer a powerful means to balance functionality and biocompatibility. We are confident that our proposed approach will help accelerate the path to useful, safe, and sustainable material science. We hope stakeholders will join this discussion and explore broader applications in the future.

■ ASSOCIATED CONTENT

Data Availability Statement

The source codes and data can be found at https://github.com/YanLabAI/De-novo-Design-Framework.

ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/envhealth.4c00088.

> Method for generating the topology of IL144 and pre-equilibrating the system (Method S1); toxicity distribution of data sets used to predict model (P) modeling—(a) Data set 1: ILs-*E. coli* MIC, (b) Data set 2: ILs-Hela cells EC_{50} (Figure S1); learning curve of the generative model (Figure S2); converting ILs'

SMILES to ASCII for synthesizing novel structures (Figure S3); experimental results of the antibacterial effect of IL144 (Figure S4); the effect of the IL144 on the survival rate of HeLa cells (Figure S5); example of stack-RNN cells with class activation features (Figure S6); NMR spectrum of the IL144 sample dissolved in DMSO (Figure S7); the average energies of both van der Waals and electrostatic interactions between IL144 anion and the membrane (Figure S8); range of hyperparameters adjusted in machine learning and deep learning (Table S1); detailed SMILES representation of 342 ILs (Table S2) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Xiliang Yan − *Institute of Environmental Research at Greater Bay Area, Key Laboratory for Water Quality and Conservation of the Pearl River Delta, Ministry of Education, Guangzhou University, Guangzhou 510006, China; College of Animal Science, South China Agricultural University, Guangzhou 510642, China;* ⓘ orcid.org/0000-0003-4173-6228; Email: yanxiliang1991@scau.edu.cn

Authors

Xiaofang Li − *Institute of Environmental Research at Greater Bay Area, Key Laboratory for Water Quality and Conservation of the Pearl River Delta, Ministry of Education, Guangzhou University, Guangzhou 510006, China*

Hanle Chen − *Institute of Environmental Research at Greater Bay Area, Key Laboratory for Water Quality and Conservation of the Pearl River Delta, Ministry of Education, Guangzhou University, Guangzhou 510006, China*

Jiachen Yan − *Institute of Environmental Research at Greater Bay Area, Key Laboratory for Water Quality and Conservation of the Pearl River Delta, Ministry of Education, Guangzhou University, Guangzhou 510006, China*

Guohong Liu − *School of Health, Guangzhou Vocational University of Science and Technology, Guangzhou 510555, China*

Chengjun Li − *Institute of Environmental Research at Greater Bay Area, Key Laboratory for Water Quality and Conservation of the Pearl River Delta, Ministry of Education, Guangzhou University, Guangzhou 510006, China*

Xiaoxia Zhou − *Institute of Environmental Research at Greater Bay Area, Key Laboratory for Water Quality and Conservation of the Pearl River Delta, Ministry of Education, Guangzhou University, Guangzhou 510006, China;* ⓘ orcid.org/0000-0001-6508-8750

Yan Wang − *College of Animal Science, South China Agricultural University, Guangzhou 510642, China*

Yinbao Wu − *College of Animal Science, South China Agricultural University, Guangzhou 510642, China;* ⓘ orcid.org/0000-0002-3235-444X

Bing Yan − *Institute of Environmental Research at Greater Bay Area, Key Laboratory for Water Quality and Conservation of the Pearl River Delta, Ministry of Education, Guangzhou University, Guangzhou 510006, China;* ⓘ orcid.org/0000-0002-7970-6764

Complete contact information is available at:
https://pubs.acs.org/10.1021/envhealth.4c00088

Author Contributions

X.L. and H.C. contributed equally to this work. X.Y. conceived this study. X.L. and H.C. wrote the paper with X.Y. and B.Y. H.C. and J.Y. developed the machine learning models. Y.W. and Y.W. analyzed the data. G.L., C.L., and X.Z. performed the wet experiments. The paper was written through the contributions of all authors. All authors have given approval for the final version of the paper.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Kaur, G.; Kumar, H.; Singla, M. Diverse applications of ionic liquids: A comprehensive review. *J. Mol. Liq.* **2022**, *351*, 118556.

(2) Hallett, J. P.; Welton, T. Room-Temperature Ionic Liquids: Solvents for Synthesis and Catalysis. 2. *Chem. Rev.* **2011**, *111*, 3508−3576.

(3) Hough, W. L.; Rogers, R. D. Ionic Liquids Then and Now: From Solvents to Materials to Active Pharmaceutical Ingredients. *Bull. Chem. Soc. Jpn.* **2007**, *80*, 2262−2269.

(4) Dai, C.; Zhang, J.; Huang, C.; Lei, Z. Ionic Liquids in Selective Oxidation: Catalysts and Solvents. *Chem. Rev.* **2017**, *117*, 6929−6983.

(5) Earle, M. J.; Seddon, K. R. Ionic liquids. Green solvents for the future. *Pure Appl. Chem.* **2000**, *72*, 1391−1398.

(6) Watanabe, M.; Thomas, M. L.; Zhang, S.; Ueno, K.; Yasuda, T.; Dokko, K. Application of Ionic Liquids to Energy Storage and Conversion Materials and Devices. *Chem. Rev.* **2017**, *117*, 7190−7239.

(7) MacFarlane, D. R.; Tachikawa, N.; Forsyth, M.; Pringle, J. M.; Howlett, P. C.; Elliott, G. D.; Davis, J. H.; Watanabe, M.; Simonf, P.; Angellg, C. A. Energy applications of ionic liquids. *Energy Environ. Sci.* **2014**, *7*, 232−250.

(8) Osada, I.; de Vries, H.; Scrosati, B.; Passerini, S. Ionic-Liquid-Based Polymer Electrolytes for Battery Applications. *Angew. Chem., Int. Ed.* **2016**, *55*, 500−513.

(9) Bi, S.; Banda, H.; Chen, M.; Niu, L.; Chen, M.; Wu, T.; Wang, J.; Wang, R.; Feng, J.; Chen, T.; Dincă, M.; Kornyshev, A. A.; Feng, G. Molecular understanding of charge storage and charging dynamics in supercapacitors with MOF electrodes and ionic liquid electrolytes. *Nat. Mater.* **2020**, *19*, 552−558.

(10) Nikfarjam, N.; Ghomi, M.; Agarwal, T.; Hassanpour, M.; Sharifi, E.; Khorsandi, D.; AliKhan, M.; Rossi, F.; Rossetti, A.; NazarzadehZare, E.; Rabiee, N.; Afshar, D.; Vosough, M.; KumarMaiti, T.; Mattoli, V.; Lichtfouse, E.; Tay, F. R.; Makvandi, P. Antimicrobial Ionic Liquid-Based Materials for Biomedical Applications. *Adv. Funct. Mater.* **2021**, *31*, 2104148.

(11) Egorova, K. S.; Gordeev, E. G.; Ananikov, V. P. Biological Activity of Ionic Liquids and Their Application in Pharmaceutics and Medicine. *Chem. Rev.* **2017**, *117*, 7132−7189.

(12) Tanner, E. E. L.; Curreri, A. M.; Balkaran, J. P. R.; Selig-Wober, N. C.; Yang, A. B.; Kendig, C.; Fluhr, M. P.; Kim, N.; Mitragotri, S. Design Principles of Ionic Liquids for Transdermal Drug Delivery. *Adv. Mater.* **2019**, *31*, 1901103.

(13) Zhang, Y.; Yuan, B.; Zhang, Y. Q.; Cao, Q. P.; Yang, C.; Li, Y.; Zhou, J. H. Biomimetic lignin/poly(ionic liquids) composite hydrogel dressing with excellent mechanical strength, self-healing properties, and reusability. *Chem. Eng. J.* **2020**, *400*, 125984.

(14) Zheng, L.; Li, J.; Yu, M.; Jia, W. B.; Duan, S.; Cao, D.; Ding, X. K.; Yu, B. R.; Zhang, X. R.; Xu, F. J. Molecular Sizes and Antibacterial Performance Relationships of Flexible Ionic Liquid Derivatives. *J. Am. Chem. Soc.* **2020**, *142*, 20257−20269.

(15) Oskarsson, A.; Wright, M. C. Ionic Liquids: New Emerging Pollutants, Similarities with Perfluorinated Alkyl Substances (PFASs). *Environ. Sci. Technol.* **2019**, *53*, 10539−10541.

(16) Wei, P.; Pan, X.; Chen, C. Y.; Li, H. Y.; Yan, X.; Li, C.; Chu, Y. H.; Yan, B. Emerging impacts of ionic liquids on eco-environmental safety and human health. *Chem. Soc. Rev.* **2021**, *50*, 13609−13627.

(17) Amde, M.; Liu, J. F.; Pang, L. Environmental Application, Fate, Effects, and Concerns of Ionic Liquids: A Review. *Environ. Sci. Technol.* **2015**, *49*, 12611−12627.

(18) Yan, J.; Yan, X.; Hu, S.; Zhu, H.; Yan, B. Comprehensive Interrogation on Acetylcholinesterase Inhibition by Ionic Liquids Using Machine Learning and Molecular Modeling. *Environ. Sci. Technol.* **2021**, *55*, 14720−14731.

(19) Stolte, S.; Arning, J.; Weber, U. B.; Matzke, M.; Stock, F.; Thiele, K.; Uerdingen, M.; Biermann, U. W.; Jastorffa, B.; Ranke, J. Anion effects on the cytotoxicity of ionic liquids. *Green Chem.* **2006**, *8*, 621−629.

(20) Cheng, C.; Ma, J. C.; Wang, J. H.; Du, Z. K.; Li, B.; Wang, J.; Gao, C.; Zhu, L. S. Toxicity comparison of three imidazolium bromide ionic liquids to soil microorganisms. *Environ. Pollut.* **2019**, *255*, 113321.

(21) Pretti, C.; Chiappe, C.; Pieraccini, D.; Gregori, M.; Abramo, F.; Monni, G.; Intorre, L. Acute toxicity of ionic liquids to the zebrafish (Danio rerio). *Green Chem.* **2006**, *8*, 238−240.

(22) Pawłowska, B.; Telesiński, A.; Biczak, R. Phytotoxicity of ionic liquids. *Chemosphere.* **2019**, *237*, 124436.

(23) Leitch, A. C.; Ibrahim, I.; Abdelghany, T. M.; Charlton, A.; Roper, C.; Vidler, D.; Palmer, J. M.; Wilson, C.; Jones, D. E.; Blain, P. G.; Wright, M. C. The methylimidazolium ionic liquid M8OI is detectable in human sera and is subject to biliary excretion in perfused human liver. *Toxicology.* **2021**, *459*, 152854.

(24) Egorova, K. S.; Ananikov, V. P. Toxicity of ionic liquids: eco(cyto)activity as complicated, but unavoidable parameter for task-specific optimization. *ChemSusChem.* **2014**, *7*, 336−360.

(25) Ekins, S.; Puhl, A. C.; Zorn, K. M.; Lane, T. R.; Russo, D. P.; Klein, J. J.; Hickey, A. J.; Clark, A. M. Exploiting machine learning for end-to-end drug discovery and development. *Nat. Mater.* **2019**, *18*, 435−441.

(26) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature.* **2018**, *559*, 547−555.

(27) Yan, X. L.; Yue, T. T.; Winkler, D. A.; Yin, Y. G.; Zhu, H.; Jiang, G. B.; Yan, B. Converting Nanotoxicity Data to Information Using Artificial Intelligence and Simulation. *Chem. Rev.* **2023**, *123*, 8575−8637.

(28) Morgan, D.; Jacobs, R. Opportunities and Challenges for Machine Learning in Materials Science. *Annu. Rev. Mate. Sci.* **2020**, *50*, 71−103.

(29) Stokes, J. M.; Yang, K.; Swanson, K.; Jin, W.; Cubillos-Ruiz, A.; Donghia, N. M.; MacNair, C. R.; French, S.; Carfrae, L. A.; Bloom-Ackermann, Z.; Tran, V. M. A Deep Learning Approach to Antibiotic Discovery. *Cell.* **2020**, *180*, 688−702.

(30) Yan, J. C.; Liu, G. H.; Chen, H. L.; Hu, S.; Wang, X. H.; Yan, B.; Yan, X. L. ILTox: A Curated Toxicity Database for Machine Learning and Design of Environmentally Friendly Ionic Liquids. *Environ. Sci. Technol. Lett..* **2023**, *10*, 983−988.

(31) Popova, M.; Isayev, O.; Tropsha, A. Deep reinforcement learning for de novo drug design. *Sci. Adv.* **2018**, *4* (7), No. eaap7885.

(32) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859−1865.

(33) Pluhackova, K.; Horner, A. Native-like membrane models of *E. coli* polar lipid extract shed light on the importance of lipid composition complexity. *BMC Biol.* **2021**, *19*, 4.

(34) Koldsø, H.; Shorthouse, D.; Hélie, J.; Sansom, M. S. P. Lipid Clustering Correlates with Membrane Curvature as Revealed by Molecular Simulations of Complex Lipid Bilayers. *PLOS Comput. Biol.* **2014**, *10*, No. e1003911.

(35) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to super-computers. *SoftwareX.* **2015**, *1−2*, 19−25.

(36) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. of Mol. Graph.* **1996**, *14*, 33−38.

(37) Bureš, F. Quaternary Ammonium Compounds: Simple in Structure, Complex in Application. *Top Curr. Chem.* **2019**, *377*, 14.

(38) Arnold, W. A.; Blum, A.; Branyan, J.; Bruton, T. A.; Carignan, C. C.; Cortopassi, G.; Datta, S.; Witt, J. D.; Doherty, A. C.; Halden, R. U.; Harari, H. Quaternary Ammonium Compounds: A Chemical Class of Emerging Concern. *Environ. Sci. Technol.* **2023**, *57*, 7645−7665.

(39) Nikfarjam, N.; Ghomi, M.; Agarwal, T.; Hassanpour, M.; Sharifi, E.; Khorsandi, D.; Khan, M. A.; Rossi, F.; Rossetti, A.; Nazarzadeh, Z. E.; Rabiee, N. Antimicrobial Ionic Liquid-Based Materials for Biomedical Applications. *Adv. Funct. Mater.* **2021**, *31*, 2104148.

(40) Elshaarawy, R. F. M.; Mokbel, W. A.; El-Sawi, E. A. Novel ammonium ionic liquids as scavengers for aromatic and heterocyclic amines: Conversion into new pharmacological agents. *J. Mol. Liq.* **2016**, *223*, 1123−1132.

(41) Shi, Y.; Zhao, Z.; Peng, K.; Gao, Y.; Wu, D.; Kim, J.; Mitragotri, S. Enhancement of Anticancer Efficacy and Tumor Penetration of Sorafenib by Ionic Liquids. *Adv. Healthc. Mater.* **2021**, *10*, 2001455.

(42) Anvari, S.; Hajfarajollah, H.; Mokhtarani, B.; Enayati, M.; Sharifi, A.; Mirzaei, M. Antibacterial and anti-adhesive properties of ionic liquids with various cationic and anionic heads toward pathogenic bacteria. *J. Mol. Liq.* **2016**, *221*, 685−690.

(43) Muñoz-Bonilla, A.; Fernández-García, M. Poly(ionic liquid)s as antimicrobial materials. *Eur. Polym. J.* **2018**, *105*, 135−149.

(44) Kuhn, B. L.; Kaminski, T. F.; Carvalho, Â. R.; Fuentefria, A. M.; Johann, B. M.; Silva, E. E.; Silveira, G. P.; Silveira, T. L.; Soares, F. A.; Zanatta, N.; Frizzo, C. P. Antimicrobial and Toxicity Evaluation of Imidazolium-Based Dicationic Ionic Liquids with Dicarboxylate Anions. *Pharmaceutics.* **2021**, *13*, 639.

(45) Aljuhani, A.; Aouad, M. R.; Rezki, N.; Aljaldy, O. A.; Al-Sodies, S. A.; Messali, M.; Ali, I. Novel pyridinium based ionic liquids with amide tethers: Microwave assisted synthesis, molecular docking and anticancer studies. *J. Mol. Liq.* **2019**, *285*, 790−802.

(46) Ranke, J.; Stolte, S.; Störmann, R.; Arning, J.; Jastorff, B. Design of Sustainable Chemical ProductsThe Example of Ionic Liquids. *Chem. Rev.* **2007**, *107*, 2183−2206.

(47) Katritzky, A. R.; Kuanar, M.; Slavov, S.; Hall, C. D.; Karelson, M.; Kahn, I.; Dobchev, D. A. Quantitative Correlation of Physical and Chemical Properties with Chemical Structure: Utility for Prediction. *Chem. Rev.* **2010**, *110*, 5714−5789.

(48) Pedro, S. N.; Freire, R. C. S.; Silvestre, A. J. D.; Freire, M. G. The Role of Ionic Liquids in the Pharmaceutical Field: An Overview of Relevant Applications. *Int. J. Mol. Sci.* **2020**, *21*, 8298.

(49) Bender, A.; Glen, R. C. Molecular similarity: a key technique in molecular informatics. *Org. Biomol. Chem.* **2004**, *2*, 3204−3218.

(50) Ertl, P.; Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. Cheminformatics.* **2009**, *1*, 8.

(51) Bonnet, P. Is chemical synthetic accessibility computationally predictable for drug and lead-like molecules? A comparative assessment between medicinal and computational chemists. *Eur. J. Med. Chem.* **2012**, *54*, 679−689.

(52) Ertl, P.; Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. Cheminform.* **2009**, *1*, 1−11.

(53) Carvalho, P. J.; Ventura, S. P.; Batista, M. L.; Schröder, B.; Gonçalves, F.; Esperança, J.; Mutelet, F.; Coutinho, J. A. Understanding the impact of the central atom on the ionic liquid behavior: Phosphonium vs ammonium cations. *J. Chem. Phys.* **2014**, *140*, 064505.

(54) Matzke, M.; Stolte, S.; Thiele, K.; Juffernholz, T.; Arning, J.; Ranke, J.; Welz-Biermann, U.; Jastorff, B. The influence of anion species on the toxicity of 1-alkyl-3-methylimidazolium ionic liquids observed in an (eco)toxicological test battery. *Green Chem.* **2007**, *9*, 1198−1207.

(55) Kumari, P.; Pillai, V. V. S.; Benedetto, A. Mechanisms of action of ionic liquids on living cells: the state of the art. *Biophys Rev.* **2020**, *12*, 1187−1215.

(56) Zhang, Y. B.; Liu, X. Y.; Wang, Y. F.; Jiang, P. P.; Quek, S. Y. Antibacterial activity and mechanism of cinnamon essential oil against Escherichia coli and Staphylococcus aureus. *Food Control.* **2016**, *59*, 282−289.

(57) Zhang, L.; Jiang, Y.; Ding, Y.; et al. Investigation into the antibacterial behavior of suspensions of ZnO nanoparticles (ZnO nanofluids). *J. Nanopart. Res.* **2007**, *9*, 479−489.

(58) Renero-Lecuna, C.; Iturrioz-Rodriguez, N.; Gonzalez-Lavado, E.; Padin-Gonzalez, E.; Navarro-Palomares, E.; Valdivia-Fernandez, L.; Garcia-Hevia, L.; Fanarraga, M. L.; Gonzalez-Legarreta, L. Effect of Size, Shape, and Composition on the Interaction of Different Nanomaterials with HeLa Cells. *Journal of Nanomaterials.* **2019**, *2019*, 1.

(59) Choudhuri, I.; Khanra, K.; Pariya, P.; et al. Structural Characterization of an Exopolysaccharide Isolated from Enterococcus faecalis, and Study on its Antioxidant Activity, and Cytotoxicity Against HeLa Cells. *Curr. Microbiol.* **2020**, *77*, 3125−3135.