Generative discovery of de novo chemical designs using diffusion modeling and transformer deep neural networks with application to deep eutectic solvents

Applied Physics Letters

RESEARCH ARTICLE | JUNE 06 2023

Generative discovery of *de novo* chemical designs using diffusion modeling and transformer deep neural networks with application to deep eutectic solvents

Special Collection: Accelerate Materials Discovery and Phenomena

Rachel K. Luu ; Marcin Wysokowski ; Markus J. Buehler

Appl. Phys. Lett. 122, 234103 (2023)
https://doi.org/10.1063/5.0155890

AIP Publishing

Generative discovery of *de novo* chemical designs using diffusion modeling and transformer deep neural networks with application to deep eutectic solvents

Cite as: Appl. Phys. Lett. **122**, 234103 (2023); doi: 10.1063/5.0155890
Submitted: 24 April 2023 · Accepted: 18 May 2023 ·
Published Online: 6 June 2023

Rachel K. Luu,[1,2] Marcin Wysokowski,[1,3] and Markus J. Buehler[1,4,a)]

AFFILIATIONS

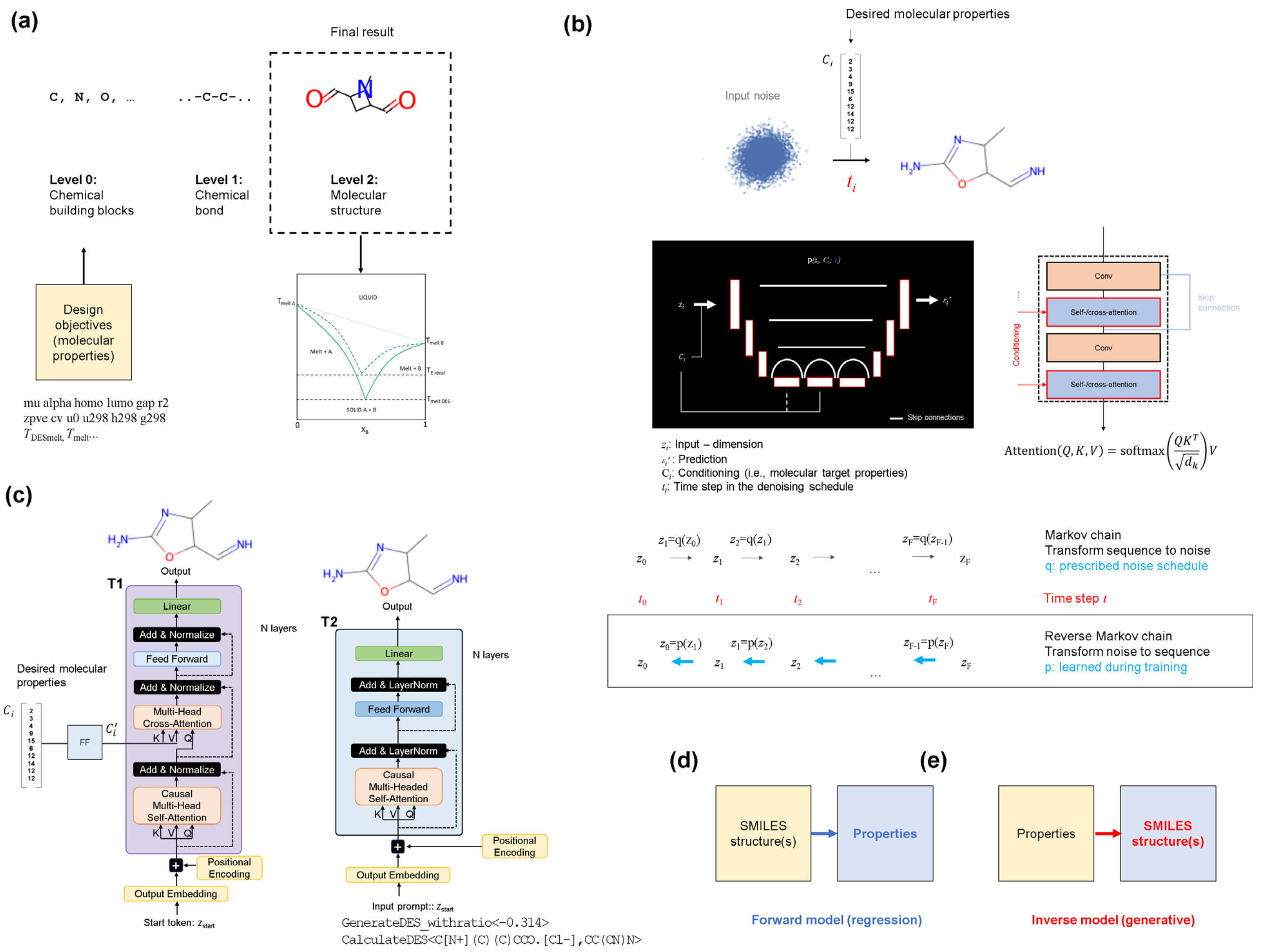
[1] Laboratory for Atomistic and Molecular Mechanics (LAMM), Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, Massachusetts 02139, USA
[2] Department of Materials Science and Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, Massachusetts 02139, USA
[3] Faculty of Chemical Technology, Poznan University of Technology, Berdychowo 4, Poznan 60965, Poland
[4] Center for Computational Science and Engineering, Schwarzman College of Computing, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, Massachusetts 02139, USA

Note: This paper is part of the APL Special Collection on Accelerate Materials Discovery and Phenomena.
[a)] **Author to whom correspondence should be addressed:** mbuehler@MIT.EDU

ABSTRACT

We report a series of deep learning models to solve complex forward and inverse design problems in molecular modeling and design. Using both diffusion models inspired by nonequilibrium thermodynamics and attention-based transformer architectures, we demonstrate a flexible framework to capture complex chemical structures. First trained on the Quantum Machines 9 (QM9) dataset and a series of quantum mechanical properties (e.g., homo, lumo, free energy, and heat capacity), we then generalize the model to study and design key properties of deep eutectic solvents (DESs). In addition to separate forward and inverse models, we also report an integrated fully prompt-based multi-task generative pretrained transformer model that solves multiple forward, inverse design, and prediction tasks, flexibly and within one model. We show that the multi-task generative model has the overall best performance and allows for flexible integration of multiple objectives, within one model, and for distinct chemistries, suggesting that synergies emerge during training of this large language model. Trained jointly in tasks related to the QM9 dataset and DESs, the model can predict various quantum mechanical properties and critical properties to achieve deep eutectic solvent behavior. Several combinations of DESs are proposed based on this framework.

Generative chemistry is an emerging frontier in materials discovery and has been applied to proteins,[1–4] organic molecules, inorganics, drug design,[5] bioactive materials,[6] solid-state materials,[7] and architected materials,[8,9] among others. Figure 1(a) shows an overview of the approach implemented, generating molecular structures from chemical building blocks, atoms. Three distinct neural network architectures are used here, a diffusion model with self-/cross-attention [Fig. 1(b)] and two transformer architectures [Fig. 1(c)]. A variety of tasks are implemented, broadly grouped into forward predictions, Fig. 1(d) [take a chemical structure written in simplified molecular-input line-entry system (SMILES; a form of textual representation that uses grammatical rules to encode information about the bonds and atoms of a molecule and allows for complex structural chemistry to be described in simple 1D text encodings)[10,11] and predict its properties] and inverse design tasks, Fig. 1(e) (take design conditions and predict candidate SMILES molecular structures). While the diffusion models and the transformer models are each trained separately for each task, the generative pretrained transformer model is fully prompt-based and capable of solving multiple tasks in one model (see Table S1 for an overview and Fig. S1 for sample results from the design processes).


(a)
C, N, O, ...
..-C-C-..
Final result
Level 0: Chemical building blocks
Level 1: Chemical bond
Level 2: Molecular structure
Design objectives (molecular properties)
mu alpha homo lumo gap r2
zpve cv u0 u298 h298 g298
T_{DESmelt}, T_{melt}...
LIQUID
Melt + A
Melt + B
SOLID A + B
$T_{\mathrm{melt\,A}}$
$T_{\mathrm{melt\,B}}$
$T_{\mathrm{E\,ideal}}$
$T_{\mathrm{melt\,DES}}$
0
1
x_B
(b)
Desired molecular properties
Input noise
C_i
t_i
$p(z_i, C_i; t_i)$
z_i
ε_i'
Skip connections
z_i: Input – dimension
ε_i': Prediction
C_i: Conditioning (i.e., molecular target properties)
t_i: Time step in the denoising schedule
Conv
Self-/cross-attention
Conditioning
skip connection
$\mathrm{Attention}(Q,K,V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$
z_1=q(z_0)
z_2=q(z_1)
z_F=q(z_{F-1})
z_0 z_1 z_2 ... z_F
t_0 t_1 t_2 t_F
Markov chain
Transform sequence to noise
q: prescribed noise schedule
Time step t
z_0=p(z_1)
z_1=p(z_2)
z_{F-1}=p(z_F)
Reverse Markov chain
Transform noise to sequence
p: learned during training
(c)
T1
Output
Linear
Add & Normalize
Feed Forward
Multi-Head Cross-Attention
Causal Multi-Head Self-Attention
K V Q
N layers
Desired molecular properties
C_i
FF
C_i'
Positional Encoding
Output Embedding
Start token: z_{start}
T2
Add & LayerNorm
Causal Multi-Headed Self-Attention
Input prompt:: z_{start}
(d)
(e)
SMILES structure(s)
Properties
Forward model (regression)
Inverse model (generative)

```
GenerateDES_withratio<-0.314>
CalculateDES<C[N+](C)(C)CCO.[Cl-],CC(CN)N>
```

FIG. 1. Approach for generating molecular structures using generative deep learning. (a) Overview of the approach used, including sample design objectives. (b) diffusion model, (c) the autoregressive transformer model, realized in several versions [T1, autoregressive transformer model where the design objectives are implemented via cross-attention, T1′ (see Fig. S7), and T2, an autoregressive transformer model where tasks are implemented via various input prompts]. Panels (d) and (e) show the tasks solved using the models.

Details on the models, training and inference process, and other key information are included in the supplementary material "S1. Materials and Methods." Here, we summarize the key components of the three architectures.

The *diffusion model* [Fig. 1(b)] implements a thermodynamics-inspired denoising process by which a noisy starting signal is transformed into the solution via a deep neural network U-net architecture.[12–14] The U-net architecture used in the diffusion model features 1D convolutional layers mixed with self-/cross-attention layers. The convolutional layers capture hierarchical patterns, and self-attention captures long-range relationships in the signal. The denoising process is conditioned using cross-attention mechanisms on a set of parameters. This results in an iterative procedure, $z_{i-1} = z_i - \varepsilon_i'$, where denoising happens by calculating the noise to be removed, ε_i', using the deep neural network p (which defines a reverse Markov chain operator), where

$$\varepsilon_i' = \mathrm{p}(z_i,\ C_i;\ t_i), \tag{1}$$

where z_i is the noisy signal at step i, C_i is the conditioning used, and t_i is the time step. The diffusion model predicts tokens as one-hot encoding, from which we then sample the token with the highest probability after denoising is complete.

The *attention-based transformer model* [Fig. 1(c), left] is implemented in several variants.[15,16] Model T1 is an autoregressive decoder-only architecture that produces solutions iteratively from a start token during inference and using cross-attention with the conditioning features. The key mathematical operation is the masked attention mechanism,[15,16] defined as

$$\mathrm{Attention}(Q, K, V; M) = \mathrm{softmax}\left(\frac{QK^T + M}{\sqrt{d_k}}\right)V. \tag{2}$$

Here, Q, K, and V are inputs to the attention layer (all the same in self-attention and Q=input, and K = V=conditioning in cross-attention). The softmax function is used here so that the attention mechanism can focus on different parts of the input sequence while considering the masking of irrelevant elements, and to ensure causality. As such, the function computes the attention scores, similar to calculating the probability or distribution of different energy states of a system. Moreover, it can be viewed as a generator function to produce edge weights of a directed graph defined by the attention mechanism, describing how strongly different elements in the input sequence interact with each other. In the inverse model, causal masking using a triangular masking matrix M is used in the self-attention step so that the model can only attend to tokens to previous tokens. We use gumbel softmax sampling during inference, which allows us to tune the creativity of the model (a certain level of noise, defined by the sampling temperature T, is added to the predicted logit distributions, from which we then sample the predicted token).[17,18] In the forward transformer model, an encoder-only strategy is used to relate the input (tokenized and embedded molecular structure) to the output, with self-attention, realizing model T1′ (Fig. S7).

In the multi-task transformer model T2 [Fig. 1(c), right], no cross-attention is necessary. All conditioning and distinction of various tasks is provided directly by the input prompt (which is first fed into the model and the model then continues the sequence to provide the answer). All code is developed in PyTorch,[19] and training is performed using an Adam optimizer.[20]

We use two datasets (for details on tokenization, etc., see supplementary material *S1.1 Tokenizer and Datasets*). The first is Quantum Machines 9 (QM9), a dataset[21] of 133 885 molecules, including their SMILES text encodings and their quantum chemical properties (QM9 is a quantum chemistry dataset consisting of 133 885 molecules; composed of the molecules' SMILES codes and its 12 associated quantum mechanical properties: dipole moment, polarizability, highest occupied molecular orbital, lowest unoccupied molecular orbital, energy gap, expectation value ⟨R2⟩, zero point energy, internal energy, internal energy at 298.15 K, enthalpy at 298.15 K, v at 298.15 K, and heat capacity at 298.15 K). Second, a smaller dataset curated for this study is used to capture properties of deep eutectic solvents (DESs). DESs are innovative mixtures comprising at least two components that can spontaneously associate to generate an eutectic phase. DESs are characterized by their significantly lower melting points in comparison with those of their individual constituents[22–24] and in comparison with the theoretical eutectic point.[25] These mixtures conserve the properties of their components, and the eutectic behavior arises from the delicate balance between the molecular dipole moments, temperature, and hydrogen bonding interactions between the constituent species. These solvents provide a clean and sustainable medium for the processing and synthesis of advanced materials[22] and are considered as impactful solvent strategies.[26,27]

Among QM9 properties that have critical impact on the generation of DESs, we list:

(i) polarizability (essential to capture the subtle balance between multiple hydrogen bonds as well as the dynamic properties[28]),
(ii) difference in energy between these frontier molecular orbitals (HOMO–LUMO gap),[29]
(iii) dipole moment,[30] and
(iv) enthalpy.[31]

In the following, we present results produced by these models. While the diffusion model and transformer model T1 take a certain type of input (tokenized SMILES strings or numerical values of target properties), the multi-task transformer model T2 is entirely based on text input and computes solutions by providing prompts (overview of some of the prompts trained for, see Table I). The purpose of utilizing these three distinct architectures and a total of five trained models is to assess the overall best strategy to dealing with the problem at hand. We specifically hypothesize that using a multi-task integrated model T2 that can be trained simultaneously on diverse datasets, and multiple tasks, can yield certain synergies and perform better overall.

This formulation of the multi-task integrated transformer model T2 offers a much more flexible approach to various kinds of design and analysis problems. The input in the transformer model T2 is purely text, for both numerical input and output as well as SMILES codes. In this framework, we use an input

```
~Calculate<CC1=CC2CC2CC1O>,
```

which is then transformed by T2 into

`~Calculate<CC1=CC2CC2CC1O>` **/-0.889,-0.176, 0.230,0.020,-0.278,-0.342,0.300, 0.370, -0.028,-0.028,-0.028,-0.028|$**,

where the output of the model is highlighted in bold. The resulting numbers of this calculation task to obtain the 12 QM9 properties can

TABLE I. Sample prompts trained for in the multi-task transformer model T2. Additional sets of tokens are used to encapsulate various tasks and input/output boundaries (∼: start token, ⟨·⟩ encapsulate task, as well as /.| to encapsulate prediction, and $ as end token).

Prompt	Prediction
∼Calculate<CC1 = CC2CC2CC1O>	Calculate QM9 properties from SMILES input
∼Generate<-0.767,-0.274,0.284,-0.020, -0.332,-0.386,0.128,0.235,-0.124, -0.124,-0.124,-0.124>	Design a molecule, expressed as a SMILES output, to meet the target QM9 properties
∼GenerateDES<-0.551,-0.570>	Generate a pair of DES molecules that meet the mole ratio and T_{DESmelt} target
∼GenerateDES_withratio<-0.487>	Generate a pair of DES molecules and associated mole ratio that meet the T_{DESmelt} target
∼CalculateTmelt<C[P+] (C1 = CC=CC=C1)(C2 = CC=CC=C2) C3 = CC=CC=C3.[Br-]>	Calculate T_{melt} for a DES component

then be converted from string format into floating point numbers for further analysis.

Figure 2 shows a comparison of design objectives labeled as ground truth (GT) vs predicted values (prediction), for the mechanical properties captured in the QM9 dataset. In this analysis, we use the workflow shown in Fig. 2(a), where this analysis tests *both* the forward and inverse tasks simultaneously. Results are shown for the three architectures used, the diffusion model [Fig. 2(b), R2 = 0.92], the transformer models T1/T1′ [Fig. 2(c), R2 = 0.94], and the prompt-based transformer model T2 [Fig. 2(d), R2 = 0.97]. To complement these results, Fig. S2 shows the forward model performance alone, predicted values over GT. Figure S2(a) shows the results for the diffusion model (R2 = 0.97), and Fig. S2(b) shows the results for the transformer model T1 (R2 = 0.96). Figure S2(c) depicts results for the transformer model T2 (R2 = 0.99), for the QM9 dataset.

We now analyze a few sample structures generated by each of the three approaches (Fig. 3) and provide an in-depth discussion of the results and performance [as shown in Fig. 3(a), the results examine conditional generation of molecules and subsequent independent testing of whether or not the desired properties were achieved]. In the following analysis, only chemical designs are considered that do not exist in the training or validation set. Figure 3(b) shows results based on the diffusion model, Fig. 3(c) for the transformer model T1, and Figs. 3(c) and 3(d) for the multi-task integrated transformer model T2, each for different conditioning parameters plotting GT vs prediction using the regression model. All three models have a strong capacity to discover structures (included in neither training or test set). We find that the multi-task integrated transformer model T2 produces generally a higher fraction of non-existing molecular designs than the other two models. This, combined with the better overall performance with respect to forward and inverse tasks, and the overall greater flexibility, indicates broader advantages of this architecture over the other two.

The multi-task integrated transformer model T2 is text based and can carry out multiple tasks. For example, in the generative task, an input,

```
~Generate<-0.767,-0.274,0.284,-0.020,
-0.332,-0.386,0.128,0.235,-0.124,-0.124,
-0.124,-0.124>,
```

leads to the output (highlighted in bold),

```
~Generate<-0.767,-0.274,0.284,-0.020,
-0.332,-0.386,0.128,0.235,-0.124,-0.124,
-0.124,-0.124> /CC1NCC2C1NC2=O|$.
```

The previous examples considered *de novo* design tasks. However, in some cases, we wish to either start with an existing chemical design or solve a partial design task where we only want to redesign part of a molecule. Such tasks can be addressed quite well using generative models, especially the diffusion approach. Figures 4(a)–4(c) show structural discovery experiments using inpainting strategy, using the inverse diffusion model. Figures 4(a) and 4(b) show results, where the first three SMILES characters are given as a fixed constraint and the rest as initial solution that can change. Figure 4(c) shows the generation results for an unconstrained design, but with an initial guess [same as in panel, Fig. 4(a)]. The highest R2 score between the desired properties and the predicted properties is obtained for the case in panel (c) (R2 = 0.86), the second highest for the case in panel a (R2 = 0.85), and the worst for the results in Fig. 4(b) (R2 = 0.82). Overall, the generative method discovers molecules that are close to the target, but the best result is obtained for the unconstrained case shown in Fig. 4(c), which makes intuitive sense. The structures in Figs. 4(a) and 4(b) are generated *de novo* by the model, whereas the structure in Fig. 4(c) is seen in the dataset. This experiment shows how by using inpainting and masking we can direct the model toward discovery of new molecules that meet a specific target and interpolate between different levels of novelty.

Figures 4(d)–4(f) show similar structural discovery experiments using the autoregressive transformer model. A distinction to the diffusion model is that due to the autoregressive nature, it does not allow for inpainting experiments; and hence, we use only three initial symbols in the SMILES string to initiate generation. (These are provided to the model after the start token.) The model completes these initial design ideas and produces molecules that meet the design demand well. The R2 values for the results in Figs. 4(d)–4(f) are R2 = 0.86, R2 = 0.82, and R2 = 0.82, respectively.

We now focus on the most complex set of tasks, making forward and inverse predictions for deep eutectic solvents (DESs). As an emerging class of mixtures, discovery of new combinations of hydrogen bond acceptors (HBAs) and donors (HBDs) that achieve DESs behavior is rather expensive and time consuming in the laboratory

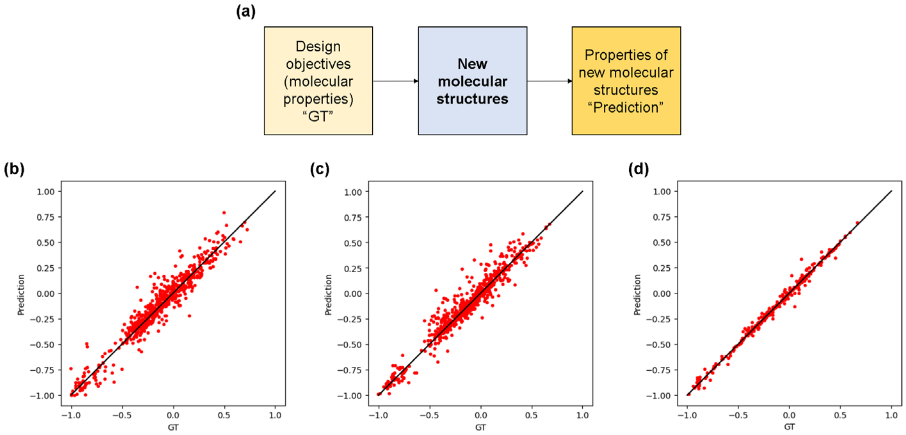


FIG. 2. Comparing design objectives, ground truth vs prediction, in the workflow as shown in panel (a) (this analysis tests both forward/inverse tasks). Results are shown for three architectures used, the diffusion model [(b) R2 = 0.92], the transformer model T1/T1′ [(c) R2 = 0.94], and the multi-task prompt-based transformer model T2 [(d) R2 = 0.97].

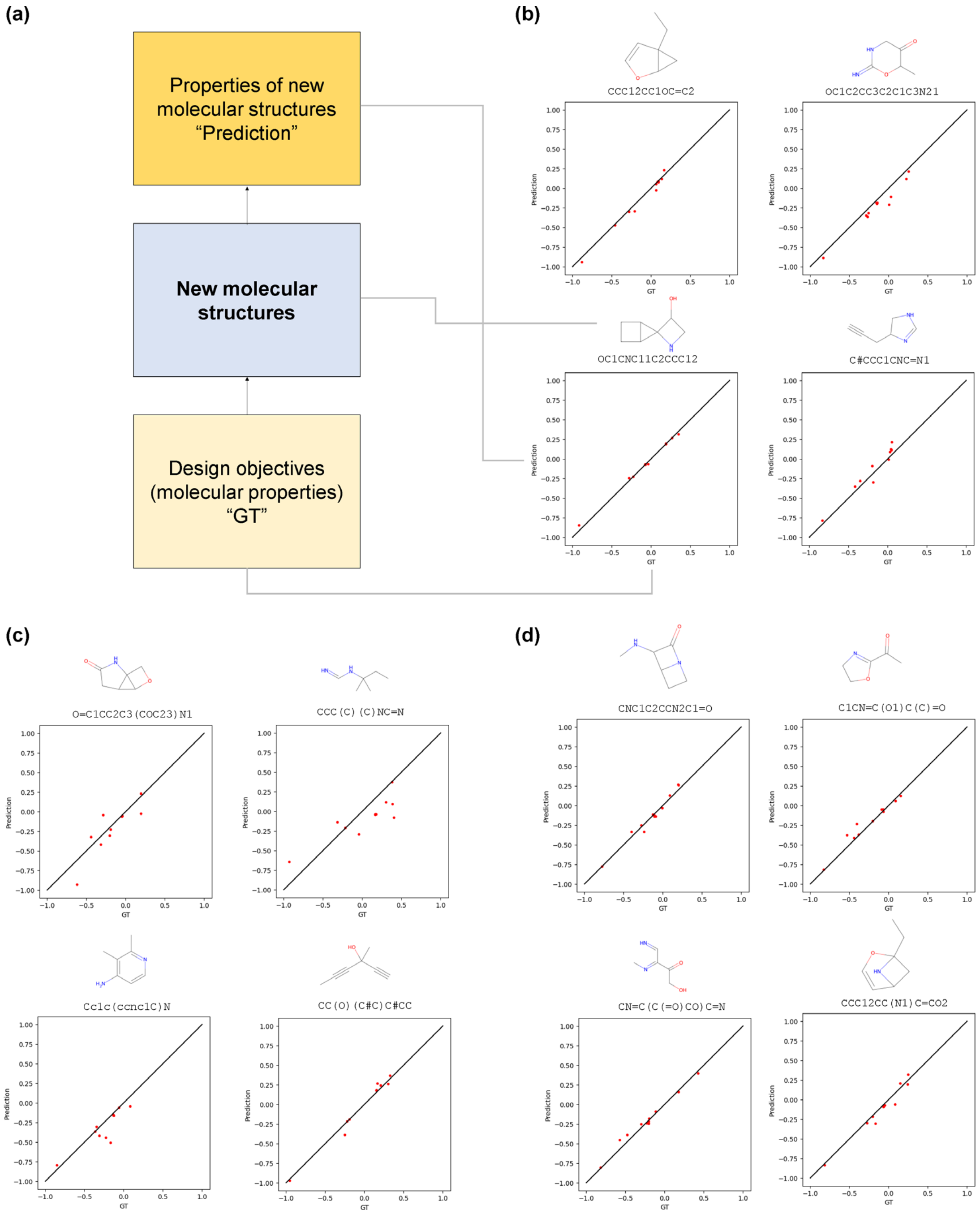


FIG. 3. Systematic analysis workflow (a) that shows that our models are generating molecular structures, nonexistent in the datasets, and exhibiting high accuracy in predicting its chemical properties when re-inserted into solving the forward problem to show consistency. Sample structures generated by the diffusion model (b), transformer model T1 (c), and transformer model T2 (d), for different conditioning parameters, plotting Ground Truth vs prediction using the regression model.

setting. Therefore, there is a growing emphasis in using computational design and machine learning algorithms as tools to support DESs discovery and predict their features,[32] including density,[33] viscosity,[34,35] and surface tension.[36] Melting temperature plays an essential role in the design of new deep eutectic solvents.[25,37] Therefore, we introduce a new dataset of DESs that contains 402 different DES compositions, where HBAs and HBDs are represented by SMILES. The dataset consists of the melting temperature of individual HBAs and HBDs, as well

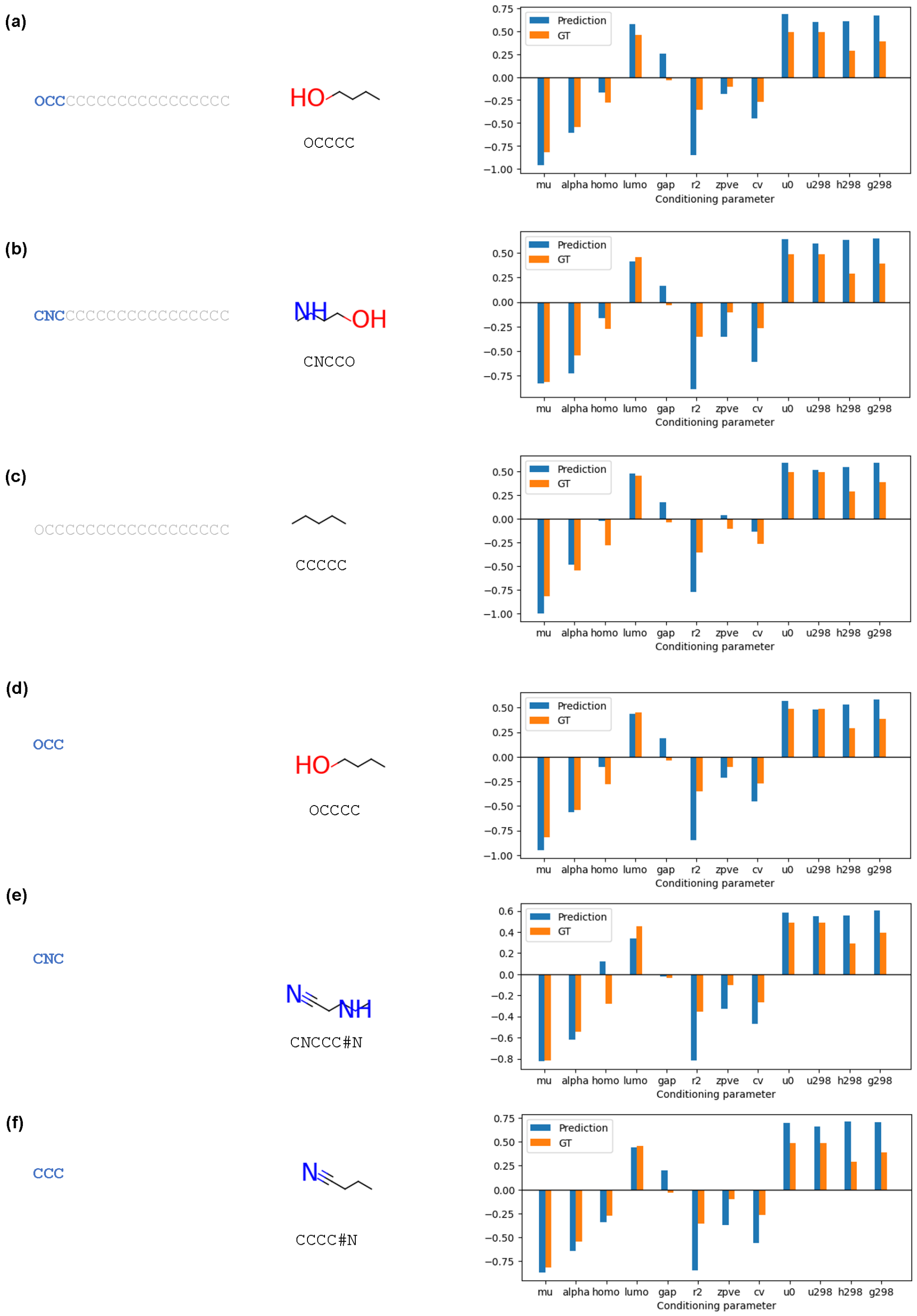


FIG. 4. Structural discovery experiments using inpainting strategy with inverse diffusion model (a)–(c). All samples are generated using 25 sampling steps, 2 resampling steps, and conditioning scale = 2.0 to increase "creativity" of the model. Panels (a) and (b) show results based on an initial structural guess, where the first three SMILES characters printed in bold blue are given as a fixed constraint (the other SMILES characters, all "C," in gray color are also provided to the model as initial guess but are changeable as the model discovers the solution). (c) Generation results for a completely unconstrained design, but with an initial guess [same as in (a)]. Panels (d)–(f) show similar experiments using the autoregressive transformer model.

as melting points of DESs mixtures in relation to their HBAs:HBDs mole ratio (see the supplementary material). The new dataset for this class of materials is much smaller than QM9 dataset (402 data points vs 130 000+). When training the models solely on the small DES dataset, we expectedly find lower performance potentially due to influential outliers and overfitting. [Figure S3 depicts performance for the forward task, Fig. S3(a) for the diffusion model (R2 = 0.59), and Fig. S3(b) for the transformer model T1′ (R2 = 0.55)]. Since the forward model does not perform well, we do not consider models trained for the inverse design task.

We propose a training strategy that can deal with such complex tasks while still providing an avenue to integrate the two datasets, both DES and QM9, for learned synergies between the various problems. Using the multi-task integrated transformer model T2, we train it against a variety of tasks (see Table I). Figure 5 shows the results obtained using the integrated multi-task transformer model T2 applied

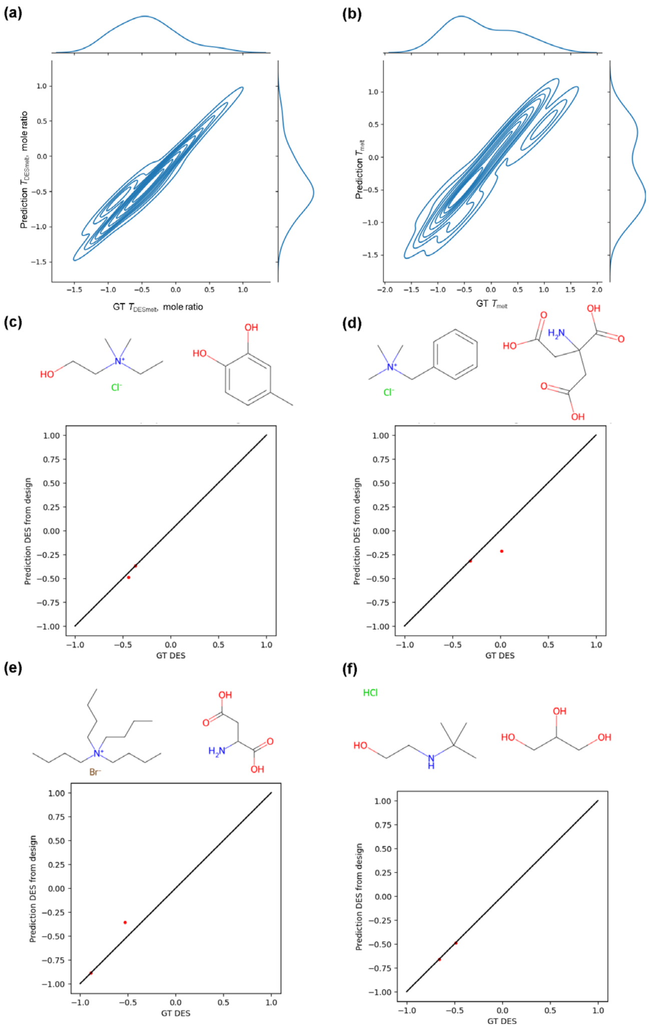


FIG. 5. Integrated multi-task transformer model T2 applied to design DESs molecular pairs and associated properties. Panels (a) and (b) show the model performance, ground truth vs prediction, trained on two different prediction tasks: (a) $T_{DESmelt}$ and mole ratio (R2 = 0.93) and (b) T_{melt}, of the individual components (R2 = 0.86). Panels (c) and (d) show two examples for results from design tasks (additional results, see supplementary material Table S2). Panels (e) and (f) document the discovery of existing designs.

to design DES molecular pairs with associated properties, trained against a combined QM9-based set of tasks (predict properties and design molecules) and DES tasks (calculate T_{melt} of individual DES components, $T_{DESmelt}$ for pairs of DES components, as well as the ratio to achieve desired $T_{DESmelt}$).

Figures 5(a) and 5(b) depict the model performance regarding DES specific prediction tasks, respectively, $T_{DESmelt}$ and ratio (R2 = 0.93) from a pair of molecules, and T_{melt}, the melting temperature of individual components (R2 = 0.86). While the predictions do not reach the same level as for the QM9 tasks shown in Fig. 2(d), the results are encouraging for machine-learning assisted DESs development. Next, we test model performance for the generative design task. Figures 5(c) and 5(d) show two example results for the design task, revealing two not-before-seen DES designs. In Fig. 5(c), $T_{DESmelt}$ and the desired mole ratio are provided, and a pair of molecules is predicted (resulting in monoethylcholine chloride and 4-methylcatechol). In Fig. 5(d), $T_{DESmelt}$ is provided, and both the pair of molecules and the mole ratio are predicted (resulting in benzyltrimethylammonium chloride and 2-aminopropane-1,2,3-tricarboxylic acid). A sample input is

```
~ GenerateDES_withratio<-0.314>,
```

which leads to

```
~GenerateDES_withratio<-0.314>
```

**/C[N+](C)(C)CC1=CC=CC=C1.[Cl-],
C(C(=O)O)C(CC(=O)O)(C(=O)O)N,0.013|$.**

The prediction is a combination of two SMILES strings and a floating-point number that describes the mole ratio.

The model was able to identify the functional groups responsible for accepting and donating the hydrogen bond and proposed new DESs composed of, for instance, monoethylcholine chloride (hydrogen bond donor count of 2) and 4-methylcatechol (hydrogen bond donor count = 2). The investigation of choline-based compounds and diols, along with aromatic alcohols,[38,39] has been extensively documented in the scientific literature. The model proposed a combination of benzyltrimethylammonium chloride (T_{melt} 239 °C) with 2-aminopropane-1,2,3-tricarboxylic acid (T_{melt} 156 °C) in molar ratio 1:1 that will result in deep depression of melting point to 33.3 °C (supplementary material Table S2). A combination of quaternary ammonium salts and carboxylic acids has been also explored in the literature.[40] For completeness, we show that the model can also rediscover already-known designs and accurately calculate DES properties for known DESs composed of tetrabutylammonium bromide and aspartic acid [Fig. 5(e)] and N,N-diethylethanolammonium chloride and glycerol [Fig. 5(f)]. It is encouraging that the model can rediscover known DES compositions. Since the dataset is extremely small, further validation of these results is needed.

The model has shown to make similar decisions as human experts in the field of DES. The results generated by the model can inspire researchers during the design phase of DES, motivating them to experimentally prepare suggested combinations of molecules and assess their T_{melt} at varying mole ratios. Subsequently, the additional experimental data could be leveraged to enhance the model's performance by adding new data to the training set.

We presented a flexible platform for materials discovery using frameworks of diffusion models and transformer architectures [Figs. 1(b) and 1(c)]. We can easily incorporate these models into a range of applications, and the use of distinct architectures offers flexible avenues. The diffusion model can easily solve inpainting problems (Fig. 4). All generative models can solve degenerate design tasks and suggest multiple candidate solutions for a given objective. The transformer models generally perform well, and the use of autoregressive approaches with multi-task training in the style of generative pretrained transformer models[41] provides the overall best performance and the highest level of flexibility [Fig. 2(c)]. The multi-task model works exceptionally well for the QM9 dataset for both forward and inverse design tasks (Figs. 2–4) and can also be applied to a new class of chemistry and associated new set of tasks (Fig. 5).

DESs are an emerging class of designer solvents in modern and sustainable chemistry for which relatively little data exist. By utilizing our newly developed dataset and applying it to our multi-task transformer model T2, we have demonstrated the impressive abilities in expediting the exploration and examination of DES properties and design. Despite using a small dataset, the model can predict diverse properties and generate new DESs compositions, like monoethylcholine halide and 4-methylcatechol. Considering the importance of DESs in toxic gas absorption, energy storage, and metal extraction, future avenues of study include the design of the DESs with multiple tailored properties like conductivity, absorption capacity, viscosity, and surface tension.

The generative models show high potential in being applied to a large variety of tasks to accelerate discovery and materials design. Another key insight is that the superb performance of the multi-task transformer model T2 has not only the best performance overall but also can be integrated efficiently with the smaller DES dataset to still yield reasonable performance. It outperforms the separately trained forward and inverse models, suggesting emergent capabilities in the language models, in general language contexts and including modeling physical phenomena.[16,42–45] This is an exciting development and offers promising future research opportunities for other systems.

See the supplementary material that includes additional details on the methods, supplementary figures, data analysis, and the DES dataset.

This material was based upon the work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 2141064. The authors acknowledge support from the Army Research Office (Nos. W911NF1920098 and W911NF2220213) as well as USDA (No. 2021-69012-35978). M.W. acknowledges the Kosciuszko Foundation and National Science Center, Poland (No. SONATA 17 2021/43/D/ST5/00853) for supporting his research stay in USA.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

M.J.B. developed the overall concept and the algorithms, designed the various deep learning models, developed the codes, oversaw the work, trained and tested the models, and drafted the paper. R.K.L. trained and tested ML models and helped to analyze results, reviewed and edited the original draft, and helped with curating the datasets. M.W. conceptualized the deep eutectic solvent work, analyzed results, reviewed, and edited the original draft, and curated the deep eutectic solvent dataset.

Rachel Kim Luu: Formal analysis (equal); Investigation (equal); Methodology (equal); Validation (equal); Visualization (equal); Writing – review & editing (equal). **Marcin Wysokowski:** Conceptualization (equal); Data curation (equal); Funding acquisition (supporting); Investigation (supporting); Validation (equal); Visualization (equal); Writing – review & editing (equal). **Markus J. Buehler:** Conceptualization (lead); Data curation (equal); Formal analysis (equal); Funding acquisition (lead); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Software (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are openly available in github at https://github.com/lamm-mit/MoleculeDiffusionTransformer, Ref. 46. Additional data, including the DES dataset, are available within the supplementary material.

REFERENCES

[1] R. Pearce and Y. Zhang, "Deep learning techniques have significantly impacted protein structure prediction and protein design," Curr. Opin. Struct. Biol. **68**, 194–207 (2021).

[2] J. Wang, H. Cao, J. Z. H. Zhang, and Y. Qi, "Computational protein design with deep learning neural networks," Sci. Rep. **8**(1), 6349 (2018).

[3] C. H. Yu, W. Chen, Y. H. Chiang, K. Guo, Z. Martin Moldes, D. L. Kaplan, and M. J. Buehler, "End-to-end deep learning model to predict and design secondary structure content of structural proteins," ACS Biomater. Sci. Eng. **8**(3), 1156–1165 (2022).

[4] B. Ni, D. L. Kaplan, and M. J. Buehler, "Generative design of de novo proteins based on secondary structure constraints using an attention-based diffusion model," Chem (published online, 2023).

[5] M. Popova, O. Isayev, and A. Tropsha, "Deep reinforcement learning for de Novo drug design," Sci. Adv. **4**, eaap7885 (2018).

[6] D. Merk, L. Friedrich, F. Grisoni, and G. Schneider, "De Novo design of bioactive small molecules by artificial intelligence," Mol. Inf. **37**(1), 1700153 (2018).

[7] B. Sanchez-Lengeling and A. Aspuru-Guzik, "Inverse molecular design using machine learning: generative models for matter engineering," Sci. **361**(6400), 360–365 (2018).

[8] A. J. Lew and M. J. Buehler, "Single-shot forward and inverse hierarchical architected materials design for nonlinear mechanical properties using an Attention-Diffusion model," Mater. Today **64**, 10 (2023).

[9] Y.-C. Hsu, Z. Yang, and M. J. Buehler, "Generative design, manufacturing, and molecular modeling of 3D architected materials based on natural language input," APL Mater. **10**(4), 041107 (2022).

[10] D. Weininger, A. Weininger, and J. L. Weininger, "SMILES. 2. Algorithm for generation of unique SMILES notation," J. Chem. Inf. Comput. Sci. **29**(2), 97–101 (1989).

[11] D. Weininger, "SMILES, a chemical language and information system: 1: Introduction to methodology and encoding rules," J. Chem. Inf. Comput. Sci. **28**(1), 31–36 (1988).

[12] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," Adv. Neural Inf. Process Syst. **33**, 6840–6851 (2020).

[13] A. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *Proceedings of the 38th International Conference on Machine Learning* (PMLR, 2021).

[14] T. Karras, M. Aittala, T. Aila, and S. Laine, "Elucidating the design space of diffusion-based generative models," arXiv:2206.00364 (2022).

[15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, *Advances in Neural Information Processing Systems* (Neural Information Processing Systems Foundation, 2017), pp. 5999–6009.

[16] Y. Hu and M. J. Buehler, "Deep language models for interpretative and predictive materials science," APL Mach. Learn. **1**(1), 010901 (2023).

[17] E. Jang, S. Gu, and B. Poole, "Categorical reparameterization with Gumbel-Softmax," in 5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings, Toulon, France, 2016.

[18] C. J. Maddison, A. Mnih, and Y. W. Teh, "The concrete distribution: A continuous relaxation of discrete random variables," in 5th International Conference

on Learning Representations, ICLR 2017 - Conference Track Proceedings, Toulon, France, 2016.

[19] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An imperative style, high-performance deep learning library," in *Proceedings of the 33rd International Conference on Neural Information Processing Systems* (Curran Associates Inc., NY, 2019), Article 721, pp. 8026–8037.

[20] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv:1412.6980 (2014).

[21] G. Chen, C.-Y. Hsieh, C.-K. Lee, B. B. Liao, J. Qiu, Q. Sun, J. Tang, and S. Zhang, "Alchemy: A quantum chemistry dataset for benchmarking AI models," arXiv:1906.09427 (2019).

[22] D. V. Wagle, H. Zhao, and G. A. Baker, "Deep eutectic solvents: Sustainable media for nanoscale and functional materials," Acc. Chem. Res. **47**(8), 2299–2308 (2014).

[23] B. B. Hansen, S. Spittle, B. Chen, D. Poe, Y. Zhang, J. M. Klein, A. Horton, L. Adhikari, T. Zelovich, B. W. Doherty, B. Gurkan, E. J. Maginn, A. Ragauskas, M. Dadmun, T. A. Zawodzinski, G. A. Baker, M. E. Tuckerman, R. F. Savinell, and J. R. Sangoro, "Deep eutectic solvents: A review of fundamentals and applications," Chem. Rev. **121**(3), 1232–1285 (2021).

[24] Q. Zhang, K. De Oliveira Vigier, S. Royer, and F. Jérôme, "Deep eutectic solvents: Syntheses, properties and applications," Chem. Soc. Rev. **41**(21), 7108–7146 (2012).

[25] D. O. Abranches and J. A. P. Coutinho, "Everything you wanted to know about deep eutectic solvents but were afraid to be told," Annu. Rev. Chem. Biomol. Eng. **14**(1), 257426919 (2023).

[26] C. Florindo, F. Lima, B. D. Ribeiro, and I. M. Marrucho, "Deep eutectic solvents: Overcoming 21st century challenges," Curr. Opin. Green Sustain. Chem. **18**, 31–36 (2019).

[27] A. Paiva, R. Craveiro, I. Aroso, M. Martins, R. L. Reis, and A. R. C. Duarte, "Natural deep eutectic solvents—Solvents for the 21st century," ACS Sustainable Chem. Eng. **2**(5), 1063–1071 (2014).

[28] K. J. Jeong, J. G. McDaniel, and A. Yethiraj, "Deep eutectic solvents: Molecular simulations with a first-principles polarizable force field," J. Phys. Chem. B **125**(26), 7177–7186 (2021).

[29] A. K. Dwamena and D. E. Raynie, "Solvatochromic parameters of deep eutectic solvents: Effect of different carboxylic acids as hydrogen bond donor," J. Chem. Eng. Data **65**(2), 640–646 (2020).

[30] P. Jahanbakhsh Bonab, A. Rastkar Ebrahimzadeh, and J. Jahanbin Sardroodi, "Insights into the interactions and dynamics of a DES formed by phenyl propionic acid and choline chloride," Sci. Rep. **11**(1), 6384 (2021).

[31] E. L. Smith, A. P. Abbott, and K. S. Ryder, "Deep eutectic solvents (DESs) and their applications," Chem. Rev. **114**(21), 11060–11082 (2014).

[32] A. Kovács, E. C. Neyts, I. Cornet, M. Wijnants, and P. Billen, "Modeling the physicochemical properties of natural deep eutectic solvents," ChemSusChem **13**(15), 3789–3804 (2020).

[33] M. Abdollahzadeh, M. Khosravi, B. Hajipour Khire Masjidi, A. Samimi Behbahan, A. Bagherzadeh, A. Shahkar, and F. Tat Shahdost, "Estimating the density of deep eutectic solvents applying supervised machine learning techniques," Sci. Rep. **12**(1), 4954 (2022).

[34] L. Y. Yu, G. P. Ren, X. J. Hou, K. J. Wu, and Y. He, "Transition state theory-inspired neural network for estimating the viscosity of deep eutectic solvents," ACS Cent. Sci. **8**(7), 983–995 (2022).

[35] D. Shi, F. Zhou, W. Mu, C. Ling, T. Mu, G. Yu, and R. Li, "Deep insights into the viscosity of deep eutectic solvents by an XGBoost-based model plus SHapley Additive exPlanation," Phys. Chem. Chem. Phys. **24**(42), 26029–26036 (2022).

[36] K. Shahbaz, F. S. Mjalli, M. A. Hashim, and I. M. AlNashef, "Prediction of the surface tension of deep eutectic solvents," Fluid Phase Equilib. **319**, 48–54 (2012).

[37] M. A. R. Martins, S. P. Pinho, and J. A. P. Coutinho, "Insights into the nature of eutectic and deep eutectic mixtures," J. Solution Chem. **48**(7), 962–982 (2019).

[38] W. Guo, Y. Hou, S. Ren, S. Tian, and W. Wu, "Formation of deep eutectic solvents by phenols and choline chloride and their physical properties," J. Chem. Eng. Data **58**, 866 (2013).

[39] R. Haghbakhsh, M. Keshtkar, A. Shariati, and S. Raeissi, "A comprehensive experimental and modeling study on CO_2 solubilities in the deep eutectic solvent based on choline chloride and butane-1,2-diol," Fluid Phase Equilib. **561**, 113535 (2022).

[40] A. P. Abbott, D. Boothby, G. Capper, D. L. Davies, and R. K. Rasheed, "Deep eutectic solvents formed between choline chloride and carboxylic acids: Versatile alternatives to ionic liquids," J. Am. Chem. Soc. **126**(29), 9142–9147 (2004).

[41] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," Preprint. https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf.

[42] S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg, H. Nori, H. Palangi, M. T. Ribeiro, and Y. Zhang, "Sparks of artificial general intelligence: Early experiments with GPT-4," arXiv:2303.12712 (2023).

[43] M. J. Buehler, "Multiscale modeling at the interface of molecular mechanics and natural language through attention neural networks," Acc. Chem. Res. **55**, 3387 (2022).

[44] M. J. Buehler, "FieldPerceiver: Domain agnostic transformer model to predict multiscale physical fields and nonlinear material properties through neural ologs," Mater. Today **57**, 9–25 (2022).

[45] Y. Hu and M. J. Buehler, "End-to-end protein normal mode frequency predictions using language and graph models and application to sonification," ACS Nano **16**(12), 20656–20670 (2022).

[46] R. K. Luu, M. Wysokowski, and M. J. Buehler, GitHub: https://github.com/lamm-mit/MoleculeDiffusionTransformer