



Article

Molecular Conditional Generation and Property Analysis of Non-Fullerene Acceptors with Deep Learning

Shi-Ping Peng , Xin-Yu Yang and Yi Zhao *

State Key Laboratory for Physical Chemistry of Solid Surfaces, Fujian Provincial Key Lab of Theoretical and Computational Chemistry, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China; psphi@stu.xmu.edu.cn (S.-P.P.); 20520181152664@stu.xmu.edu.cn (X.-Y.Y.)

* Correspondence: yizhao@xmu.edu.cn

Abstract: The proposition of non-fullerene acceptors (NFAs) in organic solar cells has made great progress in the raise of power conversion efficiency, and it also broadens the ways for searching and designing new acceptor molecules. In this work, the design of novel NFAs with required properties is performed with the conditional generative model constructed from a convolutional neural network (CNN). The temporal CNN is firstly trained to be a good string-based molecular conditional generative model to directly generate the desired molecules. The reliability of generated molecular properties is then demonstrated by a graph-based prediction model and evaluated with quantum chemical calculations. Specifically, the global attention mechanism is incorporated in the prediction model to pool the extracted information of molecular structures and provide interpretability. By combining the generative and prediction models, thousands of NFAs with required frontier molecular orbital energies are generated. The generated new molecules essentially explore the chemical space and enrich the database of transformation rules for molecular design. The conditional generation model can also be trained to generate the molecules from molecular fragments, and the contribution of molecular fragments to the properties is subsequently predicted by the prediction model.

Keywords: non-fullerene acceptors; convolutional neural networks; molecular generation model; frontier molecular orbital energies; chemical space



Citation: Peng, S.-P.; Yang, X.-Y.; Zhao, Y. Molecular Conditional Generation and Property Analysis of Non-Fullerene Acceptors with Deep Learning. *Int. J. Mol. Sci.* **2021**, *22*, 9099. <https://doi.org/10.3390/ijms22169099>

Academic Editor: Daniel Svozil

Received: 30 July 2021

Accepted: 20 August 2021

Published: 23 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Organic solar cells (OSCs) [1], one of the most promising directions for exploiting solar energy, consist of a blend of electron donor and acceptor materials. Traditionally, fullerene molecules are taken as acceptors because of their high electron affinity and mobility. For further improvement of power conversion efficiency (PCE), non-fullerene acceptors (NFAs) have recently been proposed as alternative acceptors [2]. Compared with fullerenes, NFAs have the advantages of adjustable energy level, appropriate optical absorption, good chemical stability, excellent morphology stability, and low synthesis cost, which could make them the most potential components in organic photovoltaic materials [3,4]. The NFAs have made remarkable progress in the performance of OSCs. By 2020, the PCE of Y6-based OSCs have already achieved over 18% [5]. Furthermore, in some specific applications, such as semi-transparent organic photovoltaic materials for buildings and agriculture and organic photodetectors, NFAs show promising application prospects. One of the kernel characteristics of NFAs is adjustable frontier orbital energies, especially for the highest occupied molecular orbital (HOMO) energy and the lowest unoccupied molecular orbital (LUMO) energy, which play significant roles in modulating the energy level alignments in the donor–acceptor pairs interface in order to enhance the interfacial charge separation efficiency and decrease the voltage losses [6]. However, it is still challenging to find and design the NFAs with desired orbital energies.

Machine learning is a branch of artificial intelligence. Machine learning models are trained for mining hidden insights and internal relationships from data and making classifi-

cations, predictions and generations. The past decade has seen increasingly rapid advances in the field of molecular design with machine learning [7]. As one of the most powerful tools for generating new data and approximating intractable functions, deep learning, a sub-field of machine learning, is widely applied in automatic molecular generation [8–12] and property prediction [13–19]. The sequence-based and graph-based generative models are the most widely used for molecular generation [20–26], in which the molecules are represented by simplified molecular input line entry specification (SMILES) [27] strings and molecular graphs. To constrain the models for generating molecules with desired properties, the goal-directed molecular generation becomes an increasingly important area in molecular design. Many methods, such as variational autoencoder (VAE) [28–30], generative adversarial network (GAN) [31], reinforcement learning (RL) [32,33], transfer learning (TL) [34–36], and recurrent neural network (RNN) [37] have been developed for conditional molecular generation. The VAE and GAN can build the relationship between molecular properties and the latent distribution of molecules. RL often combines with GAN to constrain the model for generating the targeted molecules. TL can fine-tune the trained models with smaller sets for generating focused molecules. RNN aggregates selected molecular descriptors along with bioactivity label as initial states for drug design. Most of the methods have RNN-based architectures in sequence models. Although many improvements for RNNs have been proposed [38–40], the constraints in parallelization and long-range context processing still remain. The architectures, such as temporal convolutional [41] or attention mechanism [42] are researched for replacing the RNNs.

The past decade has seen increasingly rapid advances in the field of molecular design [7] as well as the NFA design. The Harvard Clean Energy Project in conjunction with the IBM World Community Grid has screened over 51,000 NFAs to identify new classes and fragment ligation patterns [43]. Lee has developed the machine-learning model for predicting the efficiency of non-fullerene OSCs from frontier molecular orbital energy levels of organic materials [44]. Min et al. have introduced 565 donor/acceptor combinations in non-fullerene OSCs as a training dataset to evaluate the practicalities and the application potential of five common ML algorithms [45]. Troisi et al. have considered a database of known organic semiconductors and identified those compounds with computed electronic properties as promising NFAs [46]. Wang et al. have used machine learning to design NFAs for P3HT-based OSCs [47]. Compared with those works, our present work focuses on designing NFAs for given properties and analyzing the structure–property relationships of molecular fragments.

In our previous work [48], we used convolutional neural networks (CNNs) to construct generative and prediction models for the design and analysis of NFAs. It is demonstrated that the causal dilated CNN model is powerful for molecular generation and shows performance as good as other models at the benchmarks of Molecular Sets [26]. The purpose of the present work is to extend our previous model to a conditional generative model for generating the molecules from property distributions directly. Compared with the proposed molecular conditional generative models, the present model uses multi-property as the initial states for material design and has simpler architecture and good performance. Moreover, we further investigate the ability of the generative model for exploring the chemical space. In the model, the molecules are still represented by SMILES strings. Benefiting from the rapid progress in natural language processing, the SMILES-based models show better performance in conditional generation than graph-based models [49].

To judge whether the generated new molecules indeed have required properties, one can verify the results and analyze the structure–property relationship using the SMILES-based prediction model [48]. Many investigations show that the prediction models based on graph neural networks (GNNs) should be a natural and intuitive choice, and they also have excellent performance in the property prediction [13–19]. These GNNs include message passing neural network [13], deep tensor neural network [14], and so on. In the graph-based models, atoms and bonds are taken as nodes and edges respectively, the interactions of atoms are updated around nodes and edges. In the present work, we alternatively

choose GNN-based prediction models to analyze the frontier orbital energies of NFAs. The atomic number and the message about whether bonding or not are provided as inputs, graph attention networks (GAT) [50] are adopted as the main architecture, and the global attention mechanism as the readout phase.

With the help of the presented models, thousands of NFAs with targeted properties will be easily generated and screened out. It can be proved that the generated new molecules can expand the chemical space and provide more possible structural transformations. In addition, the conditional generation of molecules can also start with molecular fragments. The prediction model can accurately get the properties of generated molecules as well as the contributions of molecular fragments to those properties.

2. Methods and Materials

2.1. Molecular Conditional Generative Model

Figure 1a displays the molecular conditional generative model, which is based on our previous work [48]. The inputs of CNN are described by

$$z_l^0 = \text{Embed}(x_l) \parallel f_{\Phi}(y_1, \dots, y_n). \quad (1)$$

Here, y_1, \dots, y_n , the targeted properties, are transformed to vectors with a hidden size by linear transformation f_{Φ} . The x_l is the index of the l -th character in the SMILES string, and it is mapped to the vector of real numbers by an embedding layer. The \parallel represents the concatenation of two vectors. These operations are performed on all inputs $x_0 \dots, x_{L-1}$ before entering the convolutional layers. It should be noted that x_0 is the index of the addition starting character “&” and x_L is the index of the addition ending character “\n” in the SMILES string. For the generation of a string, the hidden state z_l^t of each character in the t -th layer should only relate to the states $z_0^{t-1}, \dots, z_l^{t-1}$ of previous characters. Thus, the causal convolutions are used in the model instead of normal convolutions. The diagram above in Figure 1b is the normal convolution, where two convolutional kernels are applied over one-dimensional multi-channel inputs. The input arrays in the window and each kernel are multiplied and summed by the element to get the element at the corresponding location in the output array. The zero-padding is added to two sides of the input arrays to keep the output arrays having the same length as the input arrays. The z_l^t is related to z_{l-1}^{t-1}, z_l^{t-1} and z_{l+1}^{t-1} . However, in the causal convolution shown in the diagram below in the Figure 1b, the zero-padding with size 2 is added to the left side of the input arrays, thus the output z_0^t is only related to $z_{l-2}^{t-1}, z_{l-1}^{t-1}$ and z_l^{t-1} . To incorporate longer messages of preceding characters and reduce the information redundancy, the dilated convolution is thus adopted to increase the spacing between the points of each kernel. In our model, the convolutional kernel size of CNN is taken as 3, the stride of the kernel is 1, the dilation factor is 2^{t-1} for the t -th layer, and the padding size is $(k-1)d$. After the causal convolution layer, the gated linear unit and residual networks are adopted for multichannel information transmission, and can be written as

$$Z^{t+1} = \text{GLU}(\text{Conv1d}(Z^t)) + Z^t, \quad (2)$$

where Z^t represents the inputs of the t -th layer z_0^t, \dots, z_{L-1}^t , the GLU is the gated linear unit, which splits the matrix in half along a given dimension to form a and b matrix, then computes $a \otimes \sigma(b)$, where σ is the sigmoid function and \otimes is the element-wise product between matrices. The operations are repeated in T layers as shown in the dashed box of Figure 1a. Then, the output of the last convolutional layer, z_l^T , is decoded to the vector z_l with the size of the dictionary (the classes of all characters in the dataset of SMILES strings) by a linear transformation. The z_l contains raw and unnormalized scores for each class of character. The criterion of each output is given by the cross entropy loss function

$$\text{loss}_{l+1} = -\ln\left(\frac{\exp(z_{l,x_{l+1}})}{\sum_j \exp(z_{l,j})}\right), \quad (3)$$

where the $z_{l,j}$ is the j -th value of vector z_l . The loss function aims to maximize the value $z_{l,x_{l+1}}$ in the z_l , where x_{l+1} is the index of the next character. The losses of all outputs are averaged across a string.

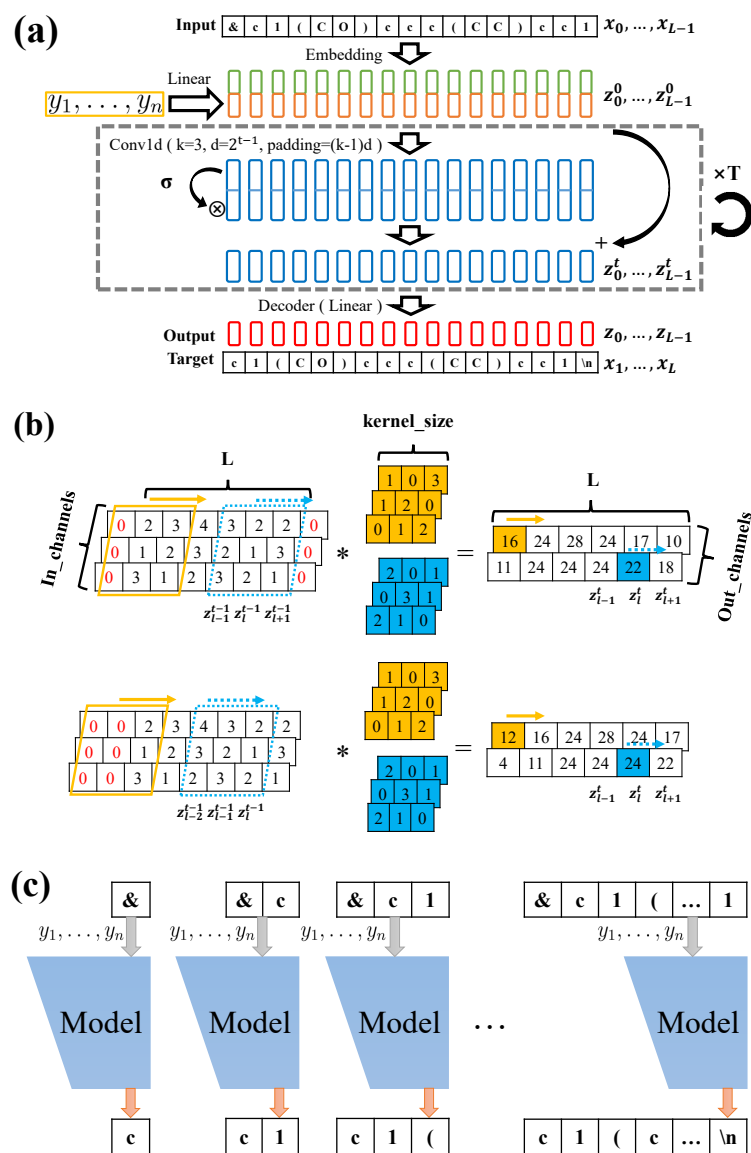


Figure 1. (a) The architecture of the generative model. (b) The normal one-dimensional multi-channel convolution and causal convolution. (c) The conditional generative process of one SMILES string.

The CNN allows parallelization over every character in a SMILES string with high efficiency during training. However, each molecular SMILES string should be generated character by character, as shown in Figure 1c. The first character is generated by taking "&" as input and the generative process is terminated as the index of "\n" is met. Between them, the character index j is sampled from the multinomial probability distribution, and the corresponding probability is given by

$$p_j = \frac{\exp(z_{l,j})}{\sum_j \exp(z_{l,j})}. \quad (4)$$

The posterior probability of the generated molecular SMILES string can be described as the product of the conditional probabilities

$$P_{\theta}(x_1, \dots, x_L | y_1, \dots, y_n) = \prod_{l=0}^{L-1} P_{\theta}(x_{l+1} | x_0, \dots, x_l, y_1, \dots, y_n). \quad (5)$$

2.2. A Graph-Based Property Prediction Model

Figure 2 exhibits the architecture of a graph-based model constructed for property prediction in the present work. The input is the atomic number c_i of each atom (node), and whether bonding or not between atoms as implicit conditions is also incorporated in the model. Then the node feature c_i is transformed to vectors h_i^0 as an initial state (hydrogen atoms are neglected in this work) by an embedding layer before entering the GAT (or other graph neural networks, the detailed architecture of GAT is provided in Appendix A). The global attention network is performed on the output h_i^T of GAT. The attention coefficient a_i of each node for output can be calculated across all nodes by the softmax function

$$a_i = \frac{\exp(f_{\Psi}(h_i^T))}{\sum_{k \in N} \exp(f_{\Psi}(h_k^T))}, \quad (6)$$

where f_{Ψ} is the linear transformation neural network. Then the output Y is achieved by aggregating messages of all nodes N

$$Y = \sum_{i \in N} a_i \odot f_{\Theta}(h_i^T), \quad (7)$$

where the hidden state h_i^T of each node is transformed to the size of output by linear transformation f_{Θ} and the \odot is element-wise multiplication. Finally, the weight of each atom to the corresponding property can be calculated by

$$w_i = \frac{a_i \odot f_{\Theta}(h_i^T)}{Y}. \quad (8)$$

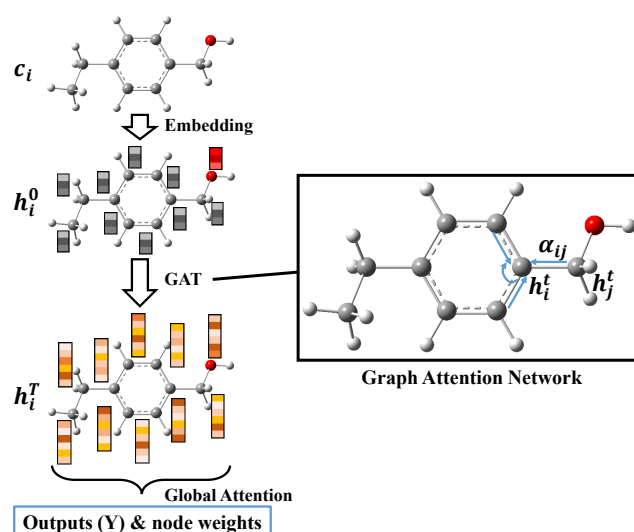


Figure 2. The architecture of the attention-based graph neural network.

2.3. Dataset and Technique Details

The library of NFAs was introduced by Lopez et al. [43]. This library has more than 51,000 molecules, built by the combinations of 12 core fragments, 43 spacer fragments, and 47 terminal fragments. The library provides the theoretical calculated and experimental calibrated data of molecular features, such as HOMO and LUMO energies, and the PCE

of solar cell with poly[N-9'-heptadecanyl-2,7-carbazole-alt-5,5-(4',7'-di-2-thienyl-2',1',3'-benzothia-diazole)] as the donor. Figure 3 shows the HOMO and LUMO distributions of the most molecules in the library. In this work, the molecular conditional generative model aims to generate desired molecules with given HOMO and LUMO energies, and the prediction model is a multi-task learning model trained for the prediction of both HOMO and LUMO energies. The filtered dataset used for the generative model contains 43,497 acceptors with the calculated HOMO energies from -7.0 eV to -5.2 eV and LUMO energies from -4.0 eV to -2.2 eV, as shown in the black frame of Figure 3. Most molecules in the dataset for the generative model have band gaps from 2 eV to 4 eV. The dataset is randomly divided into training set, validation set, and test set with sizes of 33,497, 5000, and 5000, respectively. The prediction model is trained with 50,656 acceptors whose HOMO energies locate from -7.6 eV to -4.6 eV and LUMO energies locate from -4.6 eV to -1.6 eV. Here, 40,656 molecules are used for the training set, 5000 for the validation set, and 5000 for the test set.

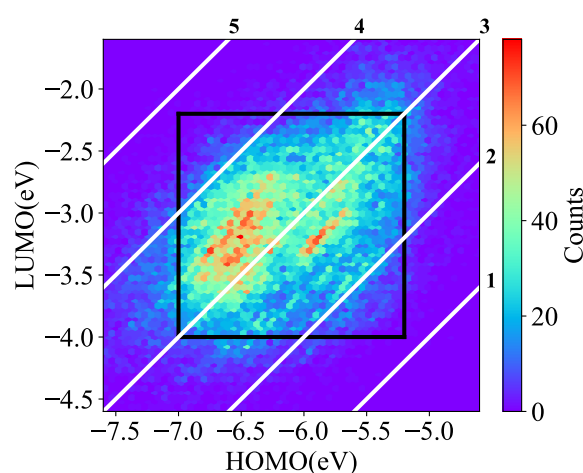


Figure 3. The distribution of NFAs for training the deep learning models (the white diagonals are the constant bandgap of the HOMO and LUMO).

The deep learning models are built with PyTorch(v1.7) [51] and DGL-LifeSci [52]. DGL-LifeSci is a python toolkit based on RDKit [53], PyTorch, and Deep Graph Library (DGL) [54]. The RDKit is an open-source tool used for related operations of cheminformatics and DGL is an open-source domain package specifically designed for researchers and application developers of GNNs. In the dataset for the generative model, the SMILES strings include 28 unique characters (exclude “&” and “\n”). In our model testing, it is found that the most suitable size of embedding layer for the best performance is 32. For the prediction model, the SMILES strings are transferred to the graph representation with the help of DGL-LifeSci. The models in this work are both trained on NVIDIA 1080Ti with a batch size of 32. Adam is taken as an optimizer with the initializing learning rate of 0.001. The model parameters with the best performance in validation set are saved during 300 epochs of training. With the trained generative model, the mean cross entropy for the test set is 0.0880. With the trained prediction model, the mean-absolute-errors/root-mean-square-errors of HOMO and LUMO energy prediction for the test set are 0.0566 eV/0.0814 eV and 0.0581 eV/0.0889 eV, respectively. More details and codes are available at <https://github.com/PSPPhi/CGEN-GPRE> (last accessed date 20 August 2021).

The new molecules generated may have new structures for designing molecules. Thus, we further use the mmpdb [55], an open-source platform, for matched molecular pair (MMP) analysis. An MMP is formed by two molecules that differ by a defined structural transformation [56], and the mmpdb can compile all possible structural transformation rules for the dataset to build the rule database and guide molecular transformation.

The quantum chemistry calculations of HOMO and LUMO energies are performed by Gaussian 16 [57] at a level of DFT/B3LYP/Def2SVP.

3. Results

3.1. Conditional Molecular Generation and Evaluation

Before applying the present generative model to NFAs generation, we first evaluate it with GuacaMol, an evaluation framework for de novo molecular design [49]. The benchmarks and performance of the model are provided in Appendix B. For the CNN-based conditional generative model trained by the dataset of NFAs, we first test its ability for generating the molecules with the properties located in the range of the training set. The 4000 targeted HOMO and LUMO energies are randomly chosen from a uniform distribution to generate the molecular SMILES strings. Then the generated SMILES strings are transformed to the molecular graph representation for the prediction of the HOMO and LUMO energies from the prediction model. Figure 4a,b displays the predicted HOMO (LUMO) energies versus targeted HOMO (LUMO) energies, in which there are 3610 reasonable molecules in the 4000 generated SMILES strings. It is seen that the targeted and predicted HOMO (LUMO) energies of the generated molecules have a relative coefficient R^2 of 0.84 (0.82). In the figures we also color the nearest neighbor molecular similarity for generated molecules. The result shows that most of the generated molecules are similar to the molecules in the dataset. To find the molecules whose predicted HOMO/LUMO energies are closest to the targeted energies, the generation processes are run 30 times (with 120,000 SMILES strings generated). Figure 4c,d show the results of 4000 generated molecules with the smallest mean absolute error (MAE) of HOMO and LUMO energies. The value of R^2 reaches to 0.99, manifesting that the molecules screened out almost meet the demand. However, the similarity shown in Figure 4c,d indicates that most of the matched molecules come from the dataset. To explore the chemical space, the generated new molecules are selected in 30 runs and 3995 best matched new molecules are screened out for 4000 desired orbital energies, and the corresponding energies are shown in Figure 4e,f. The R^2 of 0.96 indicates that the conditional molecular generation model also has an excellent performance in generating required new molecules.

Encouraged by these results, we save more than 24,000 new molecules generated here with predicted HOMO/LUMO energies in an extending library and provide it in the Supplementary Materials. This new library may be further used for exploring the chemical space. In the following section, we will demonstrate its potential applications.

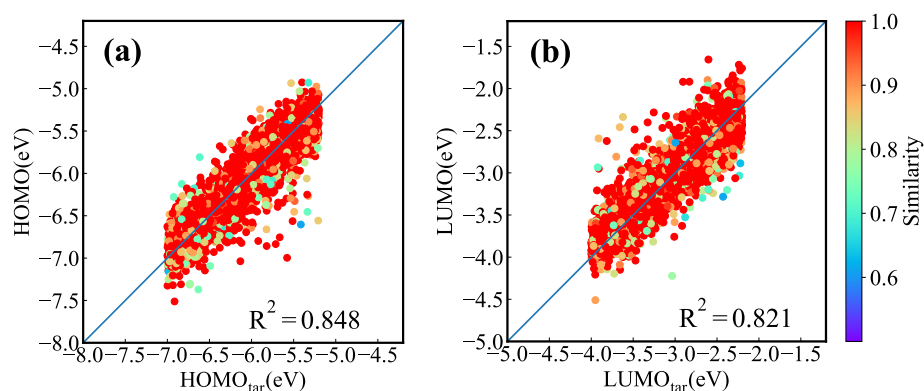


Figure 4. Cont.

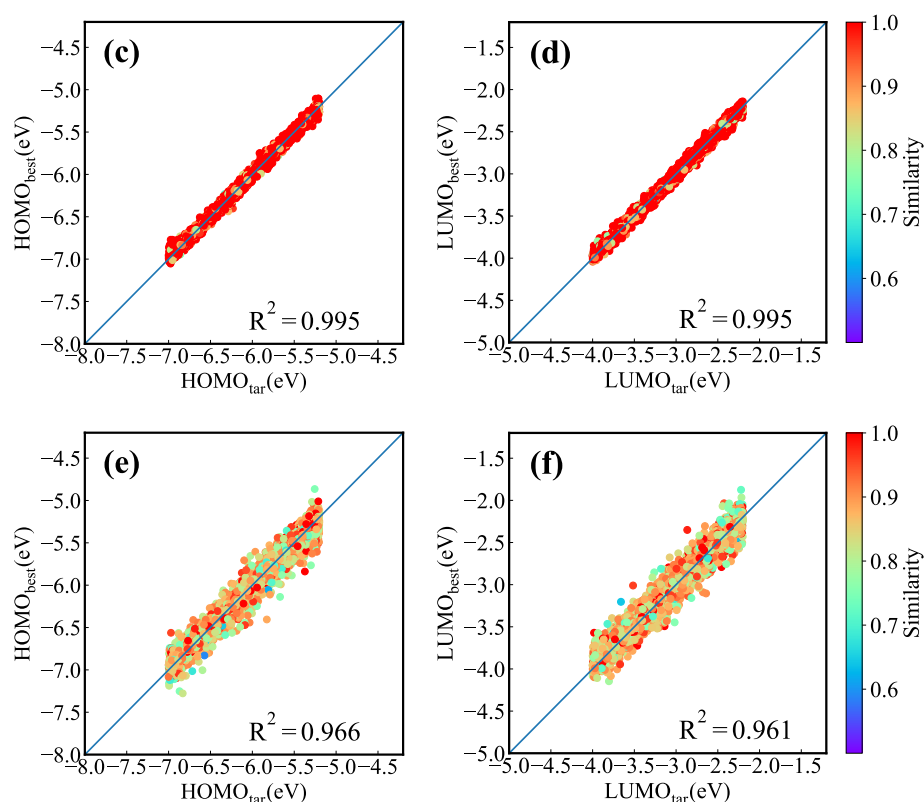


Figure 4. The performance of the molecular generation model for 4000 desired HOMO/LUMO energies. (a) The predicted HOMO energies versus targeted HOMO energies and (b) the predicted LUMO energies versus targeted LUMO energies of molecules generated in one generation process. The predicted values versus the targeted values of (c) HOMO and (d) LUMO energies of the best matched molecules screened out in 30 generation processes. The predicted values versus targeted values of (e) HOMO and (f) LUMO energies of the best matched new molecules screened out in 30 runs.

3.2. Chemical Space Exploring

We first demonstrate the ability of the conditional generative model for exploring chemical space. In the training set, the HOMO and LUMO energies are from -7.0 eV to -5.2 eV and from -4.0 eV to -2.2 eV, respectively. However, we generate molecules with the energy ranges of HOMO and LUMO larger than those in the training set, i.e., the HOMO and LUMO energies are in the range from -7.6 eV to -4.6 eV and from -4.6 eV to -1.6 eV, respectively. For a given HOMO/LUMO energy, we generate 30 SMILES strings with the conditional generative model. Figure 5a displays the MAE of predicted HOMO/LUMO and targeted HOMO/LUMO of the reasonable molecules transformed from 30 generated SMILES strings. As expected, the orbital energies of most generated molecules are close to the targeted values, but, those outside of the training set have obvious deviations, as shown on the outside of the black frame. Furthermore, in the generated molecules from 30 runs, we screen out the best-matched molecules (the predicted HOMO and LUMO energies closest to the targeted properties) and show their corresponding MAEs in Figure 5. The results indicate that the model indeed generates the molecules with required energies if these energies are not too far from the range of the training set. Specially, the generated molecules with bandgaps from 1 eV to 5 eV have low MAEs. To further reveal the ability for generating new molecules, we filter out the best matched new molecules from the molecules shown in Figure 5a, and the corresponding MAEs are shown in Figure 5c. Obviously, the MAEs of some new molecules are larger than those shown in Figure 5b.

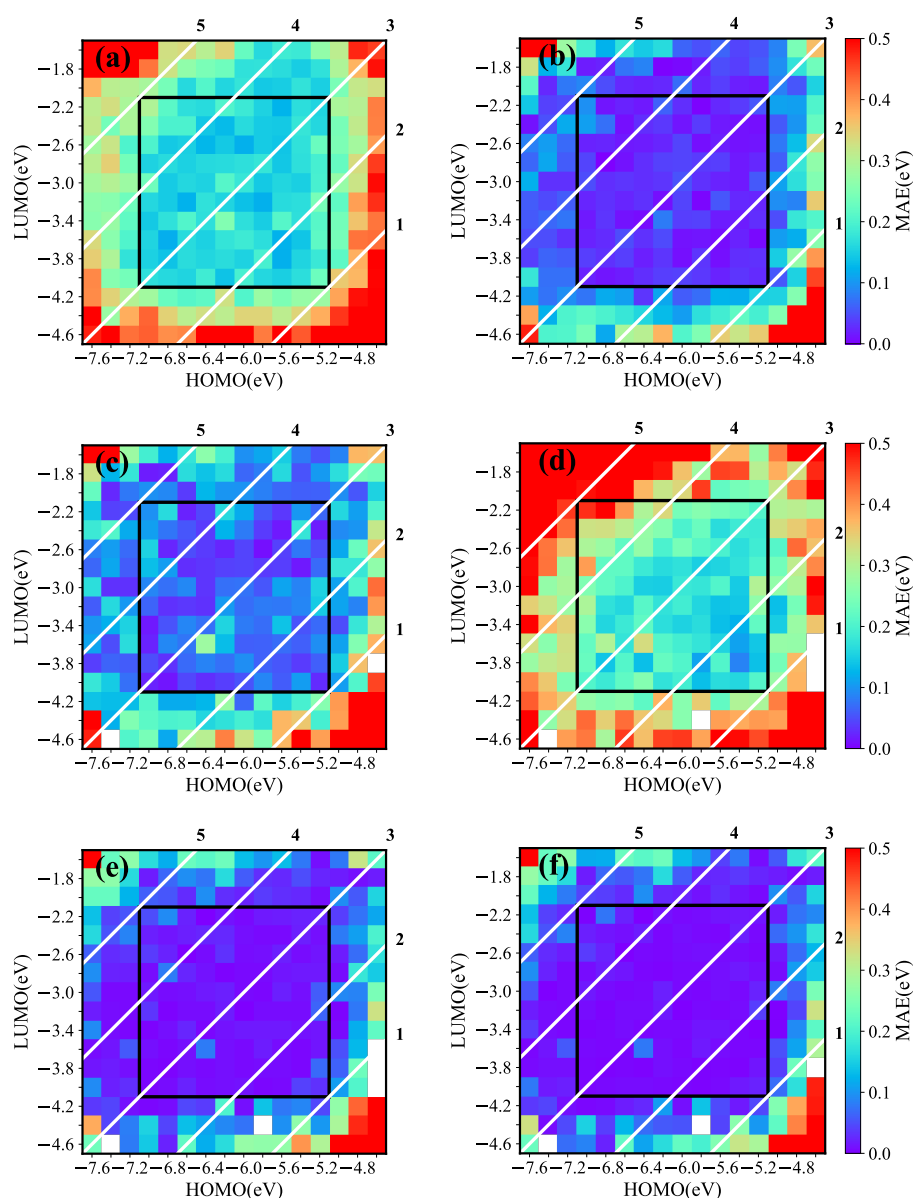


Figure 5. The performance of the conditional generative model for the desired HOMO/LUMO energies in a broad range. The heatmap shows the MAEs of predicted properties and targeted values for generated molecules (the white diagonals are the constant bandgap of the HOMO and LUMO). (a) The MAEs of molecules generated in 30 runs for desired HOMO/LUMO energies (unreasonable molecules are excluded). (b) The MAEs of the best matched molecules screened out in 30 runs. (c) The MAEs of the best matched new molecules in 30 runs. (d) The MAEs of all possible molecules transformed from molecules in (c) with the transformation rule database created by the original library. (e) The MAEs of the best matched new molecules in the transformed molecules. (f) The MAEs of the best matched new molecules transformed from molecules in (c) with the rule database created by the original library and our library.

Although new molecules may have other energies than the desired ones, they are helpful to explore the chemical space. To do so, we use “mmpdb” to build the structural transformation rule database from the original NFAs library, also the database together with our new NFAs library. Then, we apply the individual structural transformation rules from the two databases to the new molecules in Figure 5c. After that, many potential new molecules are constructed and the orbital energies of these new molecules can be predicted by the prediction model. To see the derivation of these energies to the targeted

values, we show the MAE of the molecules transformed with the original rule database in Figure 5d (the MAE from the original database together with ours have a similar behavior, which is not shown). Although the energies of transformed molecules shown in Figure 5d are distributed in a broad range, the energy errors of the best matched new molecules are smaller than the initial molecules in Figure 5c, as shown in Figure 5e, where the MAEs become comparable with those shown in Figure 5b. It is noticed that there are no transformation rules for several molecules shown by white spaces in Figure 5d,e. However, the absent transformation rules can be built up from the combination of original and our NFA libraries. For instance, the molecule at -4.6 eV (HOMO) and -3.6 eV (LUMO) appears, as shown in Figure 5f, where the best matched new molecules (marked by fork) are screened out from transformed molecules with the use of the new transformation rules database. The lower MAEs of molecules in Figure 5f show that more molecules transformed with the new database can also provide more desired molecules.

3.3. Fragment-Based Molecular Conditional Generation

The generative model can also be trained with datasets of randomized SMILES strings. With the trained model, the generation of molecules can start with fragments. Figure 6 shows the five possible molecules generated for the targeted HOMO energy of -6.5 eV and LUMO energy of -3.3 eV by the same fragment. The SMILES string of the fragment is "N#Cc1ccc2c(c1)ccs2" in its canonical form. As shown in Table 1, the strings colored in orange are the SMILES specify the last atoms with the help of Open Babel [58]. The five new molecules are generated by the trained model according to the previous characters and the targeted properties. The bonding at different geometric positions can overcome the limitations of the inherent linking relationship between fragments, and enlarge the chemical space. However, the properties of some generated molecules may differ from the targeted properties. The generated molecules can be used as the scaffolds for further decoration. Here, we choose the first molecule in Figure 6 as an example. As shown in Figure 7a, there are also five possible atoms for further decoration. Here, we specify the atom pointed by arrow 2' as an example and the SMILES string for subsequent generation is underlined. The molecule generated for the targeted HOMO energy of -6.5 eV and LUMO energy of -3.3 eV is displayed in Figure 7b. The results listed in Table 2 exhibit that the frontier orbital energies of molecule (b) are closer to the target energies.

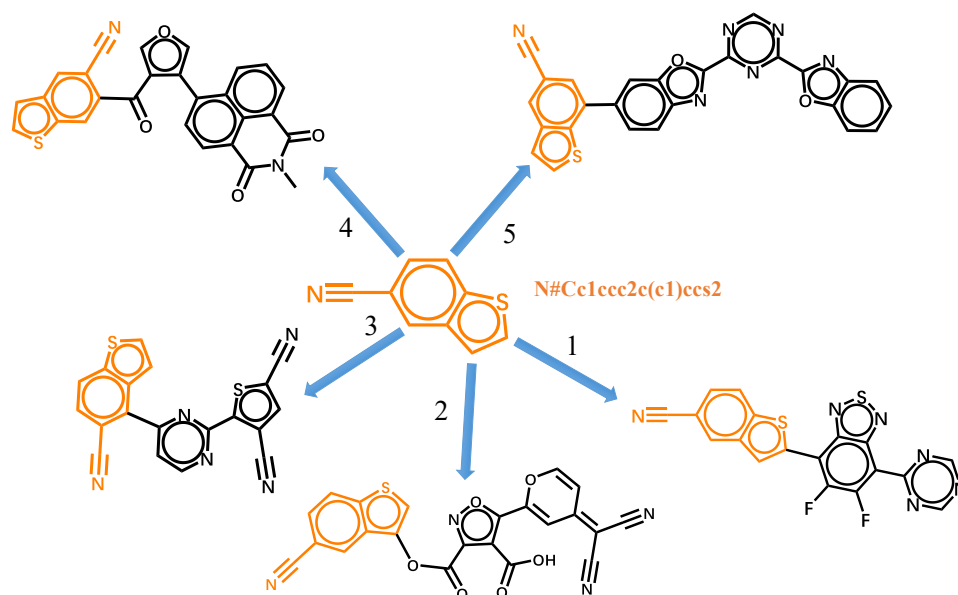
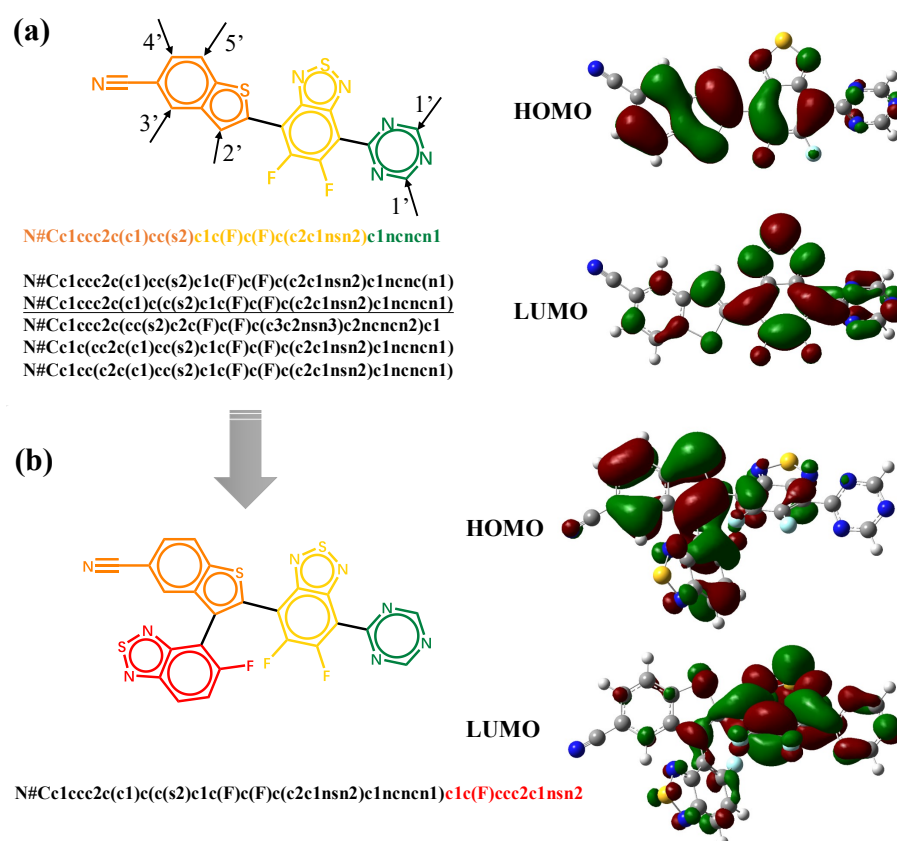


Figure 6. The five possible molecules generated from the given molecular fragment for the targeted HOMO energy of -6.5 eV and LUMO energy of -3.3 eV.

Table 1. The SMILES strings and predicted HOMO and LUMO energies of five generated molecules.

No.	SMILES	Similarity	HOMO Energy (eV)	LUMO Energy (eV)
1	N#Cc1ccc2c(c1)cc(s2)c1c(F)c(F)c(c2c1n sn2)c1ncncn1	0.75	−6.614	−3.521
2	N#Cc1ccc2c(c1)c(cs2)OC(=O)c1noc(c1 C(=O)O)C1=CC(=C(C#N)C#N)C=CO1	0.76	−6.698	−3.508
3	N#Cc1ccc2c(ccs2)c1-c1nc(-c2sc(C#N)c c2C#N)ncc1	0.78	−6.541	−3.260
4	N#Cc1c(cc2c(c1)ccs2)C(=O)c1cocc1-c1 c2c3c(ccc2)C(=O)N(C)C(=O)c3cc1	0.76	−6.468	−3.063
5	N#Cc1cc(c2c(c1)ccs2)c1ccc2nc(-c3ncnc (-c4oc5ccccc5n4)n3)oc2c1	0.81	−6.463	−2.881

The initial SMILES strings of the fragment used for generation are colored orange.

**Figure 7.** The structures of molecule (a,b) and their calculated HOMO and LUMO distributions.**Table 2.** The weighted sum of each fragment from the prediction model to the molecular orbitals for molecule (a) and (b).

Mol	Orbital	Energy_calc (eV)	Energy_pred (eV)	Red	Orange	Yellow	Green
(a)	HOMO	−6.537	−6.614		0.446	0.386	0.168
	LUMO	−3.372	−3.521		0.172	0.687	0.141
(b)	HOMO	−6.499	−6.540	0.194	0.379	0.299	0.129
	LUMO	−3.240	−3.359	0.310	0.081	0.498	0.111

It is significant to understand the molecular structure–property relationship in molecular design. In the present case, this relationship corresponds to the contributions of molecular fragments to the orbital energies, and it can be revealed by the global attention mechanism adopted in the predicted model. In Supplementary Materials, we list the results for some generated new molecules. Figure 7 displays the distributions of HOMO and LUMO obtained by Gaussian 16 calculations as the references for the verification of the predicted results. Table 2 lists the energies of HOMO and LUMO and the weighted sum of each fragment (the column header corresponds to the color of molecular fragment in the left of Figure 7) to the HOMO and LUMO energies from the predicted model. It is seen that the predicted and calculated HOMO and LUMO energies are quite close to each other. The most important fragment from the prediction model and the one from the calculation are also consistent, as shown in the table. In molecules (a) and (b), the fragment colored by orange dominantly determines HOMO energy levels, whereas the LUMO energy level-determined fragment in molecules is colored yellow. It is noticed that the prediction model can correctly predict the weights of fragments to the molecular orbital energies. The red fragment shown in Figure 7b has various weights to HOMO and LUMO. Compared with the variation of HOMO energy, the red fragment in molecule (b) introduces a larger variation of LUMO energy with a larger weight.

4. Conclusions

We have extended our previous generative model based on convolutional neural networks to incorporate the properties as conditions for generating the SMILES strings of novel molecules. The proposed conditional generative model is trained to directly generate new non-fullerene molecules with desired HOMO and LUMO energies as initial inputs. The orbital energies of generated molecules are further verified by the graph-based prediction model with the attention mechanism introduced. Together with proposed generative and prediction models trained with the library of NFAs, we have built up a new library with more than twenty thousand new non-fullerene molecules. It is found that the conditional generative model can generate new molecules with an ability to explore chemical space, and the new molecules can be used to generate the database of molecular structural transformation rules to further construct new molecules for screening. Molecular fragments can also be taken as initial conditions for generating needed molecules with a generative model trained by a dataset of randomized SMILES strings. In terms of property prediction, the prediction model predicts the HOMO and LUMO energies correctly. Furthermore, the fragment weights to these orbital energies are successfully predicted by the global attention mechanism in molecular generation processes. It should be mentioned that the HOMO and LUMO energies of new molecules in the present work are predicted by the prediction model, and the accuracy thus heavily depends on the transferability of the prediction model. To improve the accuracy, one should calibrate the orbital energies of generated new molecules with quantum chemistry calculations or experimental measurements. We believe that the architecture of CNN can be seen as an alternative to RNN in various conditional generative models and the information from interpretable models have the potential to further improve the existing models. Besides the HOMO and LUMO energies, solubility, steric hindrance, and electronic properties of NFAs, such as oscillator strength of the excited state and reorganization energy for electron transfer, are also significant for the design of NFAs. These factors can be used as additional properties for further screening if the corresponding datasets are available.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ijms22169099/s1>, Table S1: The table of generated new molecules, Table S2: The results from prediction model and quantum chemistry calculations of some generated new molecules.

Author Contributions: Conceptualization, S.-P.P. and Y.Z.; methodology, S.-P.P.; software, S.-P.P.; validation, S.-P.P., X.-Y.Y. and Y.Z.; formal analysis, S.-P.P.; investigation, S.-P.P.; resources, Y.Z.; data curation, S.-P.P. and X.-Y.Y.; writing—original draft preparation, S.-P.P.; writing—review and editing, S.-P.P.; visualization, S.-P.P.; supervision, Y.Z.; project administration, S.-P.P.; funding acquisition, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Science Foundation of China grant numbers 22033006, 21833006, and 21773191.

Data Availability Statement: The data presented in this study are available in supplementary material.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional neural network
NFA	Non-fullerene acceptor
PCE	Power conversion efficiency
HOMO	Highest occupied molecular orbital
LUMO	Lowest unoccupied molecular orbital
SMILES	Simplified molecular input line entry
VAE	Variational autoencoder
GAN	Generative adversarial network
RL	Reinforcement learning
TL	Transfer learning
RNN	Recurrent neural network
GNN	Graph neural network
GAT	Graph attention network
MMP	Matched molecular pair
MAE	Mean absolute error

Appendix A. The Graph Attention Networks

In GAT, the masked self-attention layers are performed for updating the state of each node h_i^t in t -th layer. To specify different weights of i and its nearest neighbor nodes (bonding atoms) $j \in \mathcal{N}(i)$ (including i), the softmax function is used for normalizing across all j for obtaining the attention coefficients

$$\alpha_{ij}^t = \frac{\exp(\text{LeakyReLU}((\mathbf{a}^t)^\top [W^t h_i^t \| W^t h_j^t]))}{\sum_{k \in \mathcal{N}_i} \exp(\text{LeakyReLU}((\mathbf{a}^t)^\top [W^t h_i^t \| W^t h_k^t]))} \quad (\text{A1})$$

where the h_i^t shares the same linear transformation parametrized by the weight matrix W . Then a single-layer feedforward neural network, parametrized by a weight vector \mathbf{a}^t , is applied on h_i^t and h_j^t , \top is transposition, $\|$ is concatenation, and the LeakyReLU is the activation function. Then the attention coefficients are used for pooling the states of neighborhood to obtain the updated state h_i^{t+1} of node i

$$h_i^{t+1} = \left\|_{m=1}^M \sigma \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij}^{m,t} W^m h_j^t \right) \right. \quad \text{or} \quad h_i^{t+1} = \sigma \left(\frac{1}{M} \sum_{m=1}^M \sum_{j \in \mathcal{N}_i} \alpha_{ij}^{m,t} W^m h_j^t \right). \quad (\text{A2})$$

Here, $\alpha_{ij}^{m,t}$ is the attention coefficient computed by the m -th attention mechanism and W^m is the weight matrix for linear transformation. The non-linear activation function σ and multi-head attention are used for stabilizing the model. After that, the results of K independent attention mechanisms are concatenated, except for the last layer of GAT, where averaging is applied instead of concatenation.

Appendix B. The Evaluation of CNN-Based Conditional Molecular Generation Models for Benchmarks

The performance of our model at different goal-directed benchmarks are provided in Table A1, and the table also lists the metrics for the baseline models based on genetic algorithm (GA), Monte Carlo Tree Search (MCTS), and Long Short-Term Memory (LSTM). For the similarity benchmarks, the models aim to generate molecules similar to a target that was removed from the training set. The rediscovery tasks explicitly aims to rediscover the target molecules. For the isomer benchmarks, the tasks are to generate molecules that correspond to target molecular formulas. In the median molecules benchmarks, the similarity to several molecules must be maximized simultaneously. For the multi-property objective (MPO), valsartan SMARTS, Scaffold Hop, and Decorator Hop benchmarks, each scoring function is defined as the combination of one or several functions. Here, we use the scores as the properties of our model for training and fine-tune the model with high-scoring molecules. It is demonstrated that the CNN-based conditional molecular generative model can achieve the nearly total scores as well as other baseline models. Especially, the CNN-based generative model shows better performance in similarity and isomer benchmarks.

Table A1. Results of the generative models for the goal-directed benchmarks.

Benchmark	Best of Dataset	SMILES GA	SMILES LSTM	SMILES CNN
Celecoxib rediscovery	0.505	0.732	1.000	0.844
Troglitazone rediscovery	0.419	0.515	1.000	1.000
Thiothixene rediscovery	0.456	0.598	1.000	0.687
Aripiprazole similarity	0.595	0.834	1.000	1.000
Albuterol similarity	0.719	0.907	1.000	1.000
Mestranol similarity	0.629	0.790	1.000	0.997
C ₁₁ H ₂₄	0.684	0.829	0.993	0.998
C ₉ H ₁₀ N ₂ O ₂ PF ₂ Cl	0.747	0.889	0.879	0.982
Median molecules 1	0.334	0.334	0.438	0.363
Median molecules 2	0.351	0.380	0.422	0.377
Osimertinib MPO	0.839	0.886	0.907	0.863
Fexofenadine MPO	0.817	0.931	0.959	0.976
Ranolazine MPO	0.792	0.881	0.855	0.864
Perindopril MPO	0.575	0.661	0.808	0.679
Amlodipine MPO	0.696	0.722	0.894	0.763
Sitagliptin MPO	0.509	0.689	0.669	0.586
Zaleplon MPO	0.259	0.552	0.978	0.637
Valsartan SMARTS	0.259	0.978	0.040	0.985
deco hop	0.933	0.970	0.996	0.983
scaffold hop	0.738	0.885	0.998	0.844
total	12.144	14.396	17.340	16.428

References

- Servaites, J.D.; Ratner, M.A.; Marks, T.J. Organic Solar Cells: A New Look at Traditional Models. *Energy Environ. Sci.* **2011**, *4*, 4410–4422. [[CrossRef](#)]
- Nielsen, C.B.; Holliday, S.; Chen, H.Y.; Cryer, S.J.; McCulloch, I. Non-Fullerene Electron Acceptors for Use in Organic Solar Cells. *Acc. Chem. Res.* **2015**, *48*, 2803–2812. [[CrossRef](#)] [[PubMed](#)]
- Yan, C.; Barlow, S.; Wang, Z.; Yan, H.; Jen, A.K.Y.; Marder, S.R.; Zhan, X. Non-Fullerene Acceptors for Organic Solar Cells. *Nat. Rev. Mater.* **2018**, *3*, 18003. [[CrossRef](#)]
- Zhang, J.; Tan, H.S.; Guo, X.; Facchetti, A.; Yan, H. Material Insights and Challenges for Non-Fullerene Organic Solar Cells Based on Small Molecular Acceptors. *Nat. Energy* **2018**, *3*, 720–731. [[CrossRef](#)]
- Liu, Q.; Jiang, Y.; Jin, K.; Qin, J.; Xu, J.; Li, W.; Xiong, J.; Liu, J.; Xiao, Z.; Sun, K.; et al. 18% Efficiency Organic Solar Cells. *Sci. Bull.* **2020**, *65*, 272–275. [[CrossRef](#)]

6. Zhao, Y.; Liang, W. Charge Transfer in Organic Molecules for Solar Cells: Theoretical Perspective. *Chem. Soc. Rev.* **2012**, *41*, 1075–1087. [[CrossRef](#)]
7. Mater, A.C.; Coote, M.L. Deep Learning in Chemistry. *J. Chem. Inf. Model.* **2019**, *59*, 2545–2559. [[CrossRef](#)]
8. Gupta, A.; Müller, A.T.; Huisman, B.J.H.; Fuchs, J.A.; Schneider, P.; Schneider, G. Generative Recurrent Networks for De Novo Drug Design. *Mol. Inf.* **2018**, *37*, 1700111. [[CrossRef](#)]
9. Dimitrov, T.; Kreisbeck, C.; Becker, J.S.; Aspuru-Guzik, A.; Saikin, S.K. Autonomous Molecular Design: Then and Now. *ACS Appl. Mater. Interfaces* **2019**, *11*, 24825–24836. [[CrossRef](#)]
10. Merz, K.M.; De Fabritiis, G.; Wei, G.W. Generative Models for Molecular Design. *J. Chem. Inf. Model.* **2020**, *60*, 5635–5636. [[CrossRef](#)]
11. Moosavi, S.M.; Jablonka, K.M.; Smit, B. The Role of Machine Learning in the Understanding and Design of Materials. *J. Am. Chem. Soc.* **2020**, *142*, 20273–20287. [[CrossRef](#)]
12. Thiede, L.A.; Krenn, M.; Nigam, A.; Aspuru-Guzik, A. Curiosity in Exploring Chemical Space: Intrinsic Rewards for Deep Molecular Reinforcement Learning. *arXiv* **2020**, arXiv:physics/2012.11293.
13. Gilmer, J.; Schoenholz, S.S.; Riley, P.F.; Vinyals, O.; Dahl, G.E. Neural Message Passing for Quantum Chemistry. *arXiv* **2017**, arXiv:1704.01212
14. Schütt, K.T.; Arbabzadah, F.; Chmiela, S.; Müller, K.R.; Tkatchenko, A. Quantum-Chemical Insights from Deep Tensor Neural Networks. *Nat. Commun.* **2017**, *8*, 13890. [[CrossRef](#)]
15. Schütt, K.T.; Sauceda, H.E.; Kindermans, P.J.; Tkatchenko, A.; Müller, K.R. SchNet – A Deep Learning Architecture for Molecules and Materials. *J. Chem. Phys.* **2018**, *148*, 241722. [[CrossRef](#)] [[PubMed](#)]
16. Chen, C.; Ye, W.; Zuo, Y.; Zheng, C.; Ong, S.P. Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals. *Chem. Mater.* **2019**, *31*, 3564–3572. [[CrossRef](#)]
17. Unke, O.T.; Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments and Partial Charges. *J. Chem. Theory Comput.* **2019**, *15*, 3678–3693, doi:10.1021/acs.jctc.9b00181 [[CrossRef](#)] [[PubMed](#)]
18. Korolev, V.; Mitrofanov, A.; Korotcov, A.; Tkachenko, V. Graph Convolutional Neural Networks as “General-Purpose” Property Predictors: The Universality and Limits of Applicability. *J. Chem. Inf. Model.* **2020**, *60*, 22–28. [[CrossRef](#)] [[PubMed](#)]
19. Louis, S.Y.; Zhao, Y.; Nasiri, A.; Wang, X.; Song, Y.; Liu, F.; Hu, J. Graph Convolutional Neural Networks with Global Attention for Improved Materials Property Prediction. *Phys. Chem. Chem. Phys.* **2020**, *22*, 18141–18148. [[CrossRef](#)] [[PubMed](#)]
20. Brown, N.; McKay, B.; Gilardoni, F.; Gasteiger, J. A Graph-Based Genetic Algorithm and Its Application to the Multiobjective Evolution of Median Molecules. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1079–1087. [[CrossRef](#)]
21. Ikebata, H.; Hongo, K.; Isomura, T.; Maezono, R.; Yoshida, R. Bayesian Molecular Design with a Chemical Language Model. *J. Comput.-Aided Mol. Des.* **2017**, *31*, 379–391. [[CrossRef](#)] [[PubMed](#)]
22. Blaschke, T.; Olivecrona, M.; Engkvist, O.; Bajorath, J.; Chen, H. Application of Generative Autoencoder in De Novo Molecular Design. *Mol. Inf.* **2018**, *37*, 1700123. [[CrossRef](#)] [[PubMed](#)]
23. Jensen, J.H. A Graph-Based Genetic Algorithm and Generative Model/Monte Carlo Tree Search for the Exploration of Chemical Space. *Chem. Sci.* **2019**, *10*, 3567–3572. [[CrossRef](#)]
24. Jin, W.; Barzilay, R.; Jaakkola, T. Junction Tree Variational Autoencoder for Molecular Graph Generation. *arXiv* **2019**, arXiv:1802.04364.
25. Kojima, R.; Ishida, S.; Ohta, M.; Iwata, H.; Honma, T.; Okuno, Y. kGCN: A Graph-Based Deep Learning Framework for Chemical Structures. *J. Cheminform.* **2020**, *12*, 32. [[CrossRef](#)]
26. Polykovskiy, D.; Zhebrak, A.; Sanchez-Lengeling, B.; Golovanov, S.; Tatanov, O.; Belyaev, S.; Kurbanov, R.; Artamonov, A.; Aladinskiy, V.; Veselov, M.; et al. Molecular Sets (MOSES): A Benchmarking Platform for Molecular Generation Models. *arXiv* **2020**, arXiv:1811.12823.
27. Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Model.* **1988**, *28*, 31–36. [[CrossRef](#)]
28. Gómez-Bombarelli, R.; Wei, J.N.; Duvenaud, D.; Hernández-Lobato, J.M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T.D.; Adams, R.P.; Aspuru-Guzik, A. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci.* **2018**, *4*, 268–276. [[CrossRef](#)]
29. Lim, J.; Ryu, S.; Kim, J.W.; Kim, W.Y. Molecular Generative Model Based on Conditional Variational Autoencoder for de Novo Molecular Design. *J. Cheminform.* **2018**, *10*, 31. [[CrossRef](#)]
30. Kang, S.; Cho, K. Conditional Molecular Design with Deep Generative Models. *J. Chem. Inf. Model.* **2019**, *59*, 43–52. [[CrossRef](#)]
31. Guimaraes, G.L.; Sanchez-Lengeling, B.; Outeiral, C.; Farias, P.L.C.; Aspuru-Guzik, A. Objective-Reinforced Generative Adversarial Networks (ORGAN) for Sequence Generation Models. *arXiv* **2018**, arXiv:1705.10843.
32. Olivecrona, M.; Blaschke, T.; Engkvist, O.; Chen, H. Molecular De-Novo Design through Deep Reinforcement Learning. *J. Cheminform.* **2017**, *9*, 48. [[CrossRef](#)]
33. Popova, M.; Isayev, O.; Tropsha, A. Deep Reinforcement Learning for de Novo Drug Design. *Sci. Adv.* **2018**, *4*, eaap7885. [[CrossRef](#)] [[PubMed](#)]
34. Segler, M.H.S.; Kogej, T.; Tyrchan, C.; Waller, M.P. Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks. *ACS Cent. Sci.* **2018**, *4*, 120–131. [[CrossRef](#)]

35. Amabilino, S.; Pogány, P.; Pickett, S.D.; Green, D.V.S. Guidelines for Recurrent Neural Network Transfer Learning-Based Molecular Generation of Focused Libraries. *J. Chem. Inf. Model.* **2020**, *60*, 5699–5713. [CrossRef]
36. Yasonik, J. Multiobjective de Novo Drug Design with Recurrent Neural Networks and Nondominated Sorting. *J. Cheminform.* **2020**, *12*, 14. [CrossRef]
37. Kotsias, P.C.; Arús-Pous, J.; Chen, H.; Engkvist, O.; Tyrchan, C.; Bjerrum, E.J. Direct Steering of de Novo Molecular Generation with Descriptor Conditional Recurrent Neural Networks. *Nat. Mach. Intell.* **2020**, *2*, 254–265. [CrossRef]
38. Cho, K.; van Merriënboer, B.; Bahdanau, D.; Bengio, Y. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. *arXiv* **2014**, arXiv:1409.1259.
39. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv* **2014**, arXiv:1412.3555.
40. Merity, S.; Keskar, N.S.; Socher, R. Regularizing and Optimizing LSTM Language Models. *arXiv* **2017**, arXiv:1708.02182.
41. Bai, S.; Kolter, J.Z.; Koltun, V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv* **2018**, arXiv:1803.01271.
42. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2017**, arXiv:1706.03762.
43. Lopez, S.A.; Sanchez-Lengeling, B.; de Goes Soares, J.; Aspuru-Guzik, A. Design Principles and Top Non-Fullerene Acceptor Candidates for Organic Photovoltaics. *Joule* **2017**, *1*, 857–870. [CrossRef]
44. Lee, M.H. Robust Random Forest Based Non-Fullerene Organic Solar Cells Efficiency Prediction. *Org. Electron.* **2020**, *76*, 105465. [CrossRef]
45. Wu, Y.; Guo, J.; Sun, R.; Min, J. Machine Learning for Accelerating the Discovery of High-Performance Donor/Acceptor Pairs in Non-Fullerene Organic Solar Cells. *Npj Comput. Mater.* **2020**, *6*, 120. [CrossRef]
46. Zhao, Z.W.; Omar, Ö.H.; Padula, D.; Geng, Y.; Troisi, A. Computational Identification of Novel Families of Nonfullerene Acceptors by Modification of Known Compounds. *J. Phys. Chem. Lett.* **2021**, *12*, 5009–5015. [CrossRef]
47. Mahmood, A.; Wang, J.L. A Time and Resource Efficient Machine Learning Assisted Design of Non-Fullerene Small Molecule Acceptors for P3HT-Based Organic Solar Cells and Green Solvent Selection. *J. Mater. Chem. A* **2021**, *9*, 15684–15695. [CrossRef]
48. Peng, S.P.; Zhao, Y. Convolutional Neural Networks for the Design and Analysis of Non-Fullerene Acceptors. *J. Chem. Inf. Model.* **2019**, *59*, 4993–5001. [CrossRef]
49. Brown, N.; Fiscato, M.; Segler, M.H.; Vaucher, A.C. GuacaMol: Benchmarking Models for de Novo Molecular Design. *J. Chem. Inf. Model.* **2019**, *59*, 1096–1108. [CrossRef]
50. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. *arXiv* **2018**, arXiv:1710.10903.
51. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 8026–8037
52. Li, M.; Zhou, J.; Hu, J.; Fan, W.; Zhang, Y.; Gu, Y.; Karypis, G. DGL-LifeSci: An Open-Source Toolkit for Deep Learning on Graphs in Life Science. *arXiv* **2021**, arXiv:q-bio/2106.14232.
53. RDKit. Available online: <https://www.rdkit.org/> (accessed on 4 May 2021)
54. Wang, M.; Zheng, D.; Ye, Z.; Gan, Q.; Li, M.; Song, X.; Zhou, J.; Ma, C.; Yu, L.; Gai, Y.; et al. Deep Graph Library: A Graph-Centric, Highly-Performant Package for Graph Neural Networks. *arXiv* **2020**, arXiv:1909.01315.
55. Dalke, A. Mmpdb: An Open-Source Matched Molecular Pair Platform for Large Multiproperty Data Sets. *J. Chem. Inf. Model.* **2018**, *58*, 902–910. [CrossRef] [PubMed]
56. Leach, A.G.; Jones, H.D.; Cosgrove, D.A.; Kenny, P.W.; Ruston, L.; MacFaul, P.; Wood, J.M.; Colclough, N.; Law, B. Matched Molecular Pairs as a Guide in the Optimization of Pharmaceutical Properties; a Study of Aqueous Solubility, Plasma Protein Binding and Oral Exposure. *J. Med. Chem.* **2006**, *49*, 6672–6682. [CrossRef] [PubMed]
57. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Scalmani, G.; Barone, V.; Petersson, G.A.; Nakatsuji, H.; et al. Gaussian 16 Rev. A.03. 2016
58. O’Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An Open Chemical Toolbox. *J. Cheminform.* **2011**, *3*, 33. [CrossRef] [PubMed]