

Efficient Adversarial Generation of Thermally Activated Delayed Fluorescence Molecules

Zheng Tan,* Yan Li, Ziyang Zhang, Xin Wu, Thomas Penfold, Weimei Shi, and Shiqing Yang

Cite This: *ACS Omega* 2022, 7, 18179–18188

Read Online

ACCESS |



Metrics & More

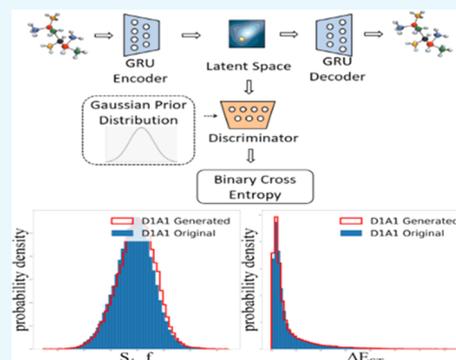


Article Recommendations



Supporting Information

ABSTRACT: Adversarial generative models are becoming an essential tool in molecular design and discovery due to their efficiency in exploring the desired chemical space with the assistance of deep learning. In this article, we introduce an integrated framework by combining the modules of algorithmic synthesis, deep prediction, adversarial generation, and fine screening for the purpose of effective design of the thermally activated delayed fluorescence (TADF) molecules that can be used in the organic light-emitting diode devices. The retrosynthetic rules are employed to algorithmically synthesize the D–A complex based on the empirically defined donor and acceptor moieties, which is followed by the high-throughput labeling and prediction with the deep neural network. The new D–A molecules are subsequently generated via the adversarial autoencoder, with the excited-state property distributions perfectly matching those of the original samples. Fine screening of the generated molecules, including the spin–orbital coupling calculation and the excited-state optimization, is eventually implemented to select the qualified TADF candidates within the novel chemical space. Further investigation shows that the created structures fully mimic the original D–A samples by maintaining a significant charge transfer characteristic, a minimal adiabatic singlet–triplet gap, and a moderate spin–orbital coupling that are desirable for the delayed fluorescence.



INTRODUCTION

Thermally activated delayed fluorescence (TADF) molecules are widely known to be highly promising materials for making high efficiency organic light-emitting diode (OLED) devices.^{1–11} In TADF, the nonemissive triplet excitons are harvested via the reverse intersystem crossing (RISC), provided the singlet–triplet gap is small enough so that the RISC can be thermally activated. These molecular systems are considered superior to conventional fluorescent and phosphorescent emitters, as they can simultaneously deliver enhanced internal quantum efficiency and cost-effective molecular design without the need for heavy atoms. However, the design of TADF molecules is challenging since it appears to require the optimization of two opposing quantities, under the approximation that the lowest singlet (S_1) and triplet (T_1) states have a charge transfer characteristic and are dominated by transitions between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO). The oscillator strength f , which corresponds to the radiative rate and therefore lifetime of S_1 , is proportional to the square of the transition dipole moment $\int \varphi_a(r)r\varphi_b(r)d^3r$, where φ_a and φ_b are the wavefunctions of the initial and final states, respectively. An efficient emission requires f to be maximized. This points toward a strong overlap of the frontier transition orbitals, i.e., HOMO and LUMO. On the other hand, the energy gap (ΔE_{ST}) between S_1 and T_1 corresponds to the exchange integral $\int \varphi_a(r_1)\varphi_b(r_2)(1/r_{12})\varphi_a(r_2)\varphi_b(r_1)-$

$d^3r_1d^3r_2$, where φ_a and φ_b are the initial and final wavefunctions, respectively, and r_{12} is the interelectronic distance. Qualified TADF candidates need a large RISC rate, which is equivalent to minimizing ΔE_{ST} and effectively reducing the overlap of frontier orbitals.

Extensive research in the past decade has been dedicated to the exploration of molecular search space to deliver emitters exhibiting strong delayed fluorescence. A large family of TADF compounds has been experimentally^{12–14} and computationally^{15,16} uncovered. Currently, most of the TADF architectures comprise electron donor and acceptor moieties. The molecular frontier orbitals, i.e., HOMO and LUMO, are deemed capable of realizing an efficient separation through the donor–acceptor dihedral twisting provided that the HOMO–LUMO transition constitutes the major transition in the first excited state, which is mostly the case in TADF. By having a small ΔE_{ST} , the donor, acceptor, and linker are carefully modulated to keep an acceptable oscillator strength and a large enough torsional angle in the meantime. To date, most of the design of new

Received: April 11, 2022

Accepted: May 11, 2022

Published: May 20, 2022



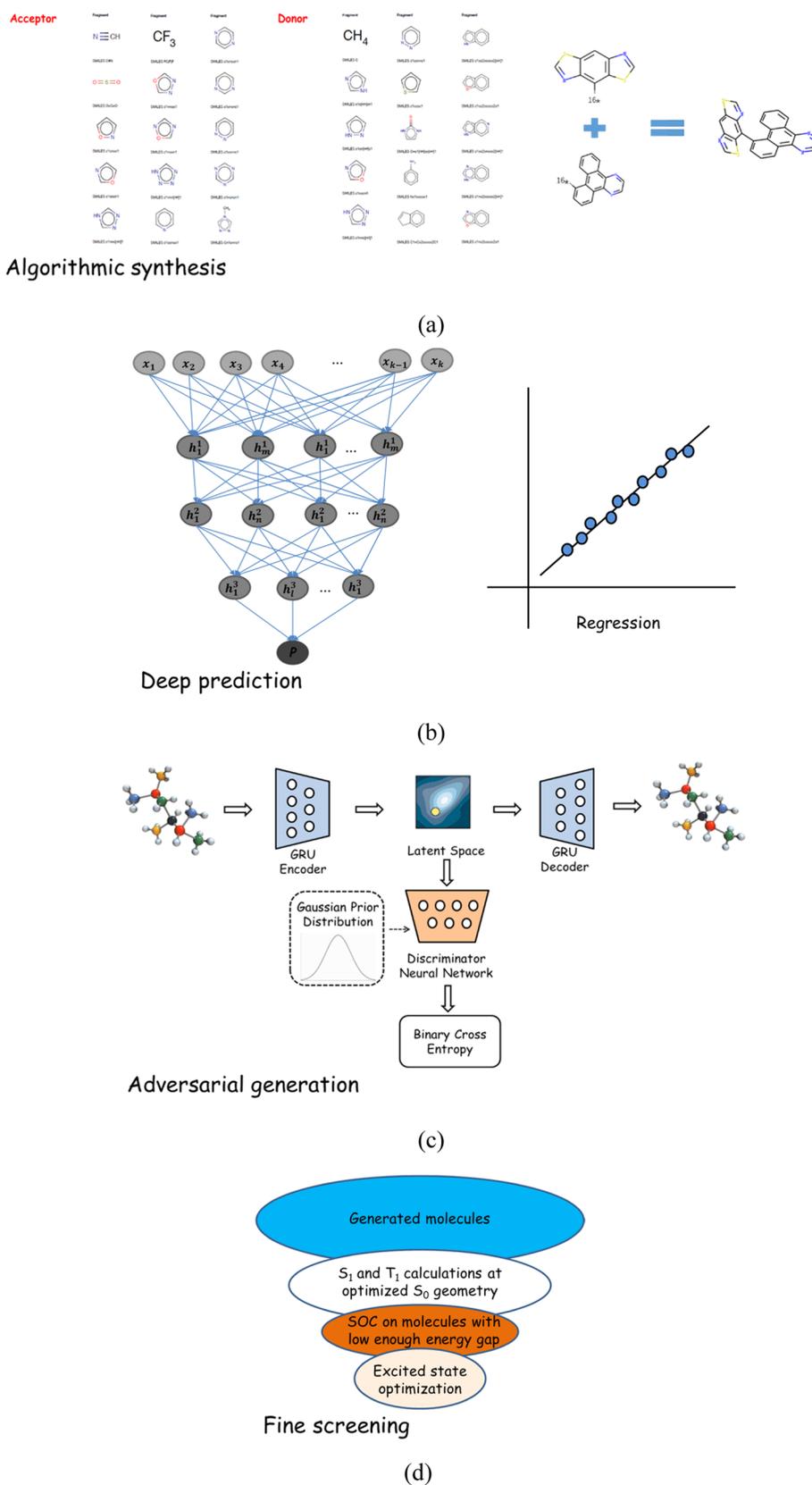


Figure 1. Integrated modules for the generation and screening of TADF molecules. (a) Algorithmic synthesis for the D–A complexes using combinatorial chemistry. (b) Deep prediction for the excited-state properties based on molecular ECFP fingerprints. (c) Adversarial generation for the TADF chemical space. (d) Fine screening of the generated molecules based on high-throughput TDDFT calculations.

materials has tended to focus on large-scale synthetic programs and trial and error. Given the size of chemical space, this

both inefficient and unlikely to lead to the development of transformative new molecules as it will be influenced by human

bias. In recent years, approaches based upon high-throughput screening and the relevant machine learning techniques have been introduced to expand the possible TADF chemical space.^{15,17,18} Gómez-Bombarelli et al.¹⁷ built a virtual library based on the donor–(bridge)_n–acceptor architecture, where time-dependent density functional theory (TDDFT) was used to label the molecular excited-state properties and deep neural network (DNN) model was employed to screen the unknown candidates. Zhao et al.¹⁸ have implemented a high-throughput virtual screening over the Cambridge Structural Database; interestingly, novel TADF emitters that go beyond the conventional donor–acceptor structure were identified leading to diversified design rules for the delayed fluorescent molecules.

Adversarial modeling is a rapidly developing field in machine learning, and it has led to a variety of great successes when applied to image generation, high-quality speech construction, and text composition. It is also very attractive to generative chemistry, since the method can learn the properties of specific real training examples and then automatically generate new synthetic entities with similar characteristics.¹⁹ Several groups in industry and academia have reported the use of adversarial models to inversely design the drug compound and optoelectronic materials with desired pharmacological and physicochemical properties^{20–26} by employing adversarial autoencoder (AAE), generative adversarial network (GAN), reinforcement learning assisted GAN, etc.

To further expand the TADF search space, this article introduces an alternative framework by integrating algorithmic synthesis, deep prediction, adversarial generation, and fine screening for the efficient computational design of molecules with preferentially excited-state properties. Given the empirically defined donor and acceptor fragments, we employ the breaking of retrosynthetically interesting chemical substructures²⁷ (BRICS) route to algorithmically synthesize 160671 D–A complexes. Part of the synthesized samples is selected for TDDFT calculations and labeled with the excited-state properties. We subsequently train a DNN model based on the labeled samples. The developed model is then used to predict the properties of generated molecules. The synthesized D–A chemical space is then fed into an adversarial autoencoder to generate molecules with similar photophysical properties. Fine screening is finally implemented to screen the generated samples for qualified TADF candidates, where extensive quantum chemical computations are carried out with the molecular excited-state geometry, spin–orbital coupling (SOC), frontier orbital transition, and adiabatic gap being examined. We find that the generated molecules can bear almost the same distributions of the excited-state properties as the original samples. Based on the structural similarity, the adversarial generation can perfectly reproduce the charge-transfer characteristics and the minimal adiabatic gap, pointing toward a feasible design philosophy for the TADF emitter.

The paper is organized as follows. The second section is dedicated to the methodology including the four modules we have for the adversarial generation. The **Results and Discussion** Section presents the empirical results and discusses the main findings of our study. A conclusion is drawn in the end.

METHODOLOGY

Algorithmic Synthesis. As depicted in Figure 1, 100 acceptors and 100 donors are first empirically defined (Tables S1 and S2 in the Supporting Information). The donors and

acceptors follow the molecular components used by Gómez-Bombarelli et al.¹⁷ The BRICS²⁷ rule is employed to construct the molecular fragment by introducing the link atoms according to the BRICS chemical environments (16 environments in total). A D–A complex is then synthesized in accordance with the retrosynthetic principle given the constructed donor and acceptor fragments. For example, the fragments shown in Figure 1a represent the 16 chemical environments (aromatic ring system), which are indicated by the link atoms at the respective reactive sites, where the combination of the two fragments leads to a synthetic D–A sample. In principle, there are a variety of combination modes for TADF: D–A, D–A–D, A–D–A, D–bridge–A, etc. For the sake of simplicity, we only build D1A1 molecules to examine how much the adversarial generation profile can mimic the original chemical space. In total, 160671 samples are algorithmically synthesized.

Deep Prediction. A total of 13594 molecules within the D1A1 chemical space are selected for the excited-state calculations. The samples are first processed with RDKit for the initial geometry optimization using the MMFF94 force field.²⁸ Ground-state geometries are reoptimized at the B3LYP/6-31G(d,p) level using Gaussian 16.²⁹ Excited-state calculations are carried out with TDDFT/B3LYP/6-31G(d,p) also using Gaussian 16. The excited-state properties at the S₀ geometry, including the first three singlet and triplet excited-state energies, the first three singlet oscillator strengths, and the first singlet–triplet splitting, are exported for the sample labeling. The 13594 labeled molecules are partitioned as 8/1/1 into training, validation, and test sets to be used for the DNN training, hyperparametrization, and generalization, respectively. The extended-connectivity fingerprints (ECFPs) are applied to represent the topological characteristics of molecules.³⁰ We employ ECFP with 2048 bits for the machine learning training input and define the iteration number as 2 during the fingerprint generation. The network topology is set as Input-2048-1024-Output, where the input is 2048, denoting the molecular fingerprint, and the output is 1, representing the regressed target. The learning rate and the maximal training epoch are set to be 0.001 and 400, respectively. We take an early stop by selecting the epoch number for generalization when the minimum of the loss function is reached in the validation set. Ten single-target DNN models are trained for each excited-state property and further employed as predictors for the property forecasting of generated samples.

Adversarial Generation. In our study, an adversarial autoencoder is chosen for molecular generation. The model basically follows the generation frameworks in MOSES³¹ using a simplified molecular-input line-entry system³² (SMILES) as input and output representations. Two neural networks, an encoder and a decoder, are utilized to map the high-dimensional string-based data onto the low-dimensional vector-based latent space. The encoder is a one-layer bidirectional gated recurrent unit (GRU) of 256 hidden dimensions, while the decoder is a three-layer GRU of 512 hidden dimensions. Note that a one-hot embedding layer is applied before the encoder to process the character sequence into digital information. Teacher forcing³³ is employed in the decoder to lower the autoencoder reconstruction loss. As shown in Figure 1c, an auxiliary discriminator network is trained to distinguish samples from a Gaussian prior distribution and the latent space. The encoder then adapts its latent space to minimize the discriminator's predictive

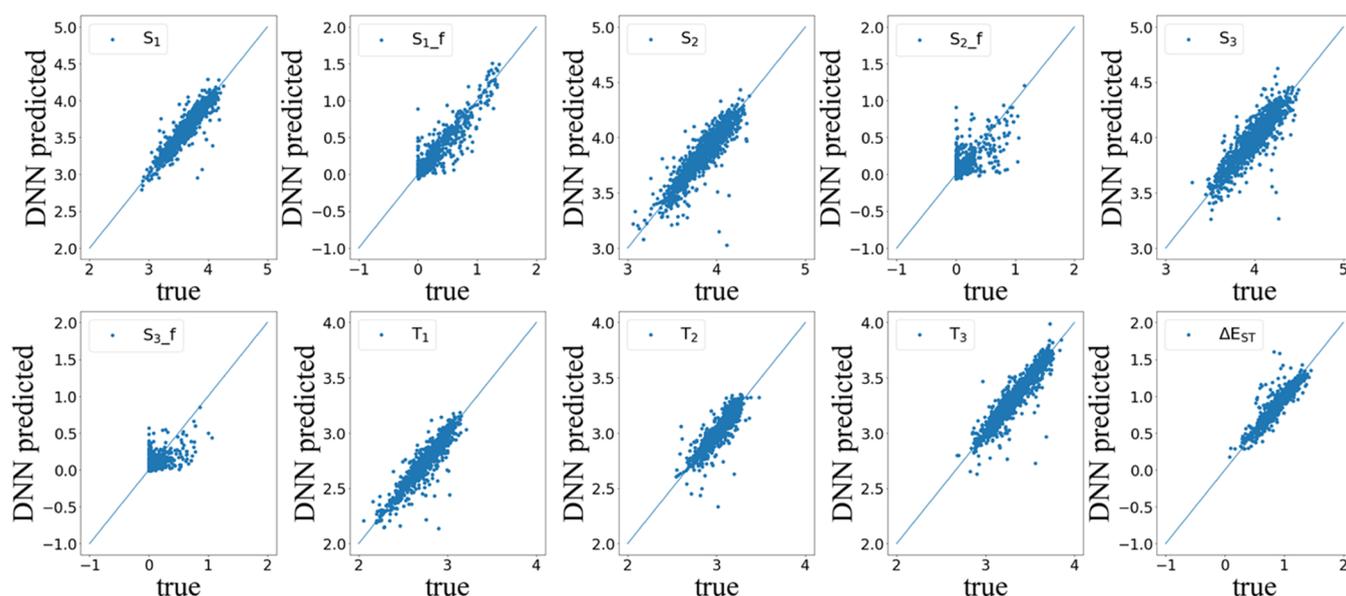


Figure 2. DNN predictions of the excited-state properties of TADF molecules. S_1 , S_2 , S_3 , T_1 , T_2 , and T_3 represent the first three singlet and triplet excited-state energies, respectively, while f denotes the corresponding oscillator strength and ΔE_{ST} the first singlet–triplet gap. Note that the true data indicate the high-throughput calculations of molecular properties. The blue line is the identity mapping. The singlet and triplet energies and the energy difference are in the unit of eV, while f is dimensionless.

accuracy. The training process oscillates between training the encoder–decoder pair and the discriminator until the latent space and the Gaussian prior are indistinguishable. The latent space is of dimension 128, and the discriminator network is a two-layer fully connected neural network with 512 and 256 nodes, respectively. The training is performed for 60 epochs, with a learning rate of 0.0005 and a batch size of 128.

Fine Screening. The generated samples from AAE are first fed into the pretrained DNN predictors with the created ECFP fingerprint to obtain the predicted excited-state properties. Part of the samples is selected according to the initial screening criterion where $\Delta E_{ST} < 0.4$ eV, and the first singlet oscillator strength $S_{1_f} > 0.02$. We set the gap criterion relatively loosely so that further screening can have more space to choose proper candidates. Three hundred thirty-five molecules from the 62 276 generated samples are screened in this step. High-throughput calculations are implemented for these 335 molecules with the same procedure as in the section of the deep prediction to acquire the TDDFT properties at the S_0 geometry. The same screening criterion ($\Delta E_{ST} < 0.4$ eV and $S_{1_f} > 0.02$) is again applied to the 335 molecules based on the “true” excited-state properties, where 67 samples are selected. The selected samples are further processed with the SOC calculations (at the S_0 geometry), which are implemented with ORCA³⁴ at the level of B3LYP/def2-SVP. Forty-seven candidates that satisfy the SOC conditions are finally picked out for excited-state optimizations, where S_1 and T_1 geometries are optimized at the B3LYP/def2-SVP level using Gaussian 16. We also perform TDDFT with the Tamm–Dancoff approximation³⁵ (TDA) in the excited-state calculations to see whether it can be helpful to reduce the problem of triplet instability.³⁶ More details in fine screening are discussed in the following sections.

RESULTS AND DISCUSSION

Excited-State Property Predictions. Figure 2 presents the DNN prediction qualities on the synthesized D1A1 test set

(10% of the 13594 labeled samples). The out-of-sample precisions are acceptable for both singlet and triplet energies, with the forecasting accuracy being gradually diminished for the higher excited state (as seen in Table S3 in Supporting Information). The observation is consistent with the previous report³⁷ where the prediction error is found to be larger as the excited state goes higher. The prediction R^2 of ΔE_{ST} is 0.8636 with the root-mean-square error (RMSE) being calculated as 0.0814 eV, which is comparable with the predictive error of the TADF rate constant given in Gómez-Bombarelli et al.¹⁷ The forecasting of singlet oscillator strengths is less accurate, especially for the higher excited states. Fortunately, the two most important properties, the singlet–triplet gap and the first singlet oscillator strength, are well fitted with R^2 above 0.85, demonstrating that the current model can capture the characteristics of delayed fluorescence.

Note that as exhibited in Figure 2, there are a bunch of “true” oscillators with strengths being zero while the predicted values are not, implying a reduced forecasting accuracy for molecules with a null f . We carefully checked the molecular geometry and found that the samples with a null f mostly possess a near-90° dihedral twisting between the donor and acceptor. The weakened prediction quality can possibly be attributed to the employed ECFP fingerprint, which is essentially two-dimensional and thus cannot incorporate the three-dimensional (3D) twisting information (known to be important to the transition probability). The analysis indicates that further improvement of the excited state forecasting accuracy may necessitate the 3D information to be encoded in the model.

To examine the feasibility of training set size in predicting the out-of-sample properties, we vary the training set size for the model training and apply the model to predict the same test set. The prediction R^2 is found to converge for both S_{1_f} and ΔE_{ST} (Figure S1 in the Supporting Information), implying that the current labeled data set is sufficient to train the model with acceptable generalizability.

Adversarial Generation Profiles. A total of 160671 original D1A1 molecules are put into AAE for the generative training. The generation performances are measured in terms of validity, novelty, and uniqueness. Validity represents the fraction of molecules within the generated samples that can pass the RDKit's sanitization check so that the atomic valency and the consistency of bonds in aromatic rings can be maintained. Novelty is the fraction of the generated molecules that are not present in the training set. Uniqueness gives rise to the percentage of the generated samples that are unique. We compute the overall generation efficiency for molecules that simultaneously fulfill the above three criteria, given different generation sizes. From Table 1, the validity and novelty do not

Table 1. Performance Metrics for the Adversarial Autoencoder Model Applied on the D1A1 Data Set: Fraction of Valid, Novel, Unique Molecules and the Overall Generation Efficiency Fulfilling the above Three Criteria Given Different Generation Sizes

generation size	validity	novelty	uniqueness	overall efficiency
1 K	0.8280	0.6490	0.9990	0.4750
10 K	0.8079	0.6587	0.9671	0.4460
100 K	0.8135	0.6503	0.8070	0.3555
200 K	0.8143	0.6537	0.7228	0.3114

vary significantly with the generation size, while the uniqueness is found to decline when more molecules are generated. The overall generation efficiency reaches a level of 0.3114 when 200 K samples are produced, which is equivalent to 62 276 molecules being effectively generated.

The internal diversity of both the original and generation sets (160671 original and 62276 generated) are computed to examine how similar the molecules are in each (see computational details in the Supporting Information). The IntDiv₁ and IntDiv₂ for the original samples are 0.8338 and 0.8233, respectively, and those for the generated samples are 0.8243 and 0.8146, respectively. The computed diversity is significant and comparable with the diversity measure in Polykovskiy et al.,³¹ demonstrating that both the synthesized and generated molecules are fully diversified.

We employ the pretrained DNN models to predict the excited-state properties for both the original and generated D1A1 molecules. The property distributions are shown in Figure 3a. Perfect matches between the original and generated distributions are found, indicating a strong structural similarity of the produced D1A1 space with the synthesized one. Both original and generated S₁ energies are roughly in the range of 2.5–4.5 eV, covering the full visible spectrum for the light emission. All of the singlet oscillator strengths tend toward a zero-side distribution. The singlet–triplet gap virtually follows a normal distribution, with a rather small proportion of molecules satisfying the gap criterion for TADF (usually 0.25 eV is the accepted upper limit³). We also inspect the synthetic accessibility (SA)³⁸ based on the fragment contributions and the molecular complexity. Both the original and generated samples bear similar distributions of SA, as seen in Figure 3b. A major peak at around 2.3 is observed, demonstrating that the D1A1 samples possess a molecular synthesizability approaching the catalog molecules.³⁸

Fine Screening of TADF Candidates. Given the generated D1A1 molecules, we further take a series of fine screening procedures to see whether the qualified TADF

candidates can be selected from the generation space. The initial screening is performed according to the predicted excited-state properties of generated samples. By applying a loose criterion ($\Delta E_{ST} < 0.4$ eV and $S_{1f} > 0.02$), 335 samples are selected from the 62276 generated molecules. The selected samples are subsequently processed with DFT calculations (at the level of B3LYP/6-31G(d,p)) to obtain the ground-state geometry, based on which the excitation energies, oscillator strengths, and the energy difference are computed at the level of TDDFT/ B3LYP/6-31G(d,p). Further screening of these 335 samples with the “true” excited-state properties leads to 67 qualified candidates that can meet the same selection criterion. We note the discrepancy in the number of molecules satisfying the criterion based on the DNN-predicted and TDDFT-calculated properties. A prediction RMSE of 0.1888 for ΔE_{ST} and 0.1053 for S_{1f} are found for the generated molecules, which is slightly higher than that obtained from the test set in the original space (0.0814 for ΔE_{ST} and 0.0964 for S_{1f}). Additional examinations show that the structural similarity in terms of ECFP is limited between the original and generated space (see Tanimoto similarity analysis in the Supporting Information), which may partially explain the lowered predictive accuracy. The observation is consistent with the findings in Popova et al.,³⁹ where the prediction and generative model are performed in a separated manner with reduced forecasting precisions being visualized for newly generated compounds. We, therefore, claim that the current exercise for property predictions within the generated space is reasonable and sufficient for the screening procedure to be performed. Considering the property predictions in other generative models,^{40,41} further enhancement of the forecasting efficiency may entail a prediction–generation model integration, which is subject to future research.

SOC calculations are performed on the S₀ geometry for those 67 generated molecules with small enough vertical gaps. Both T₁–S₁ and T₁–S₀ SOC constants are computed. We found that most candidates have T₁–S₁ coupling below 5 cm⁻¹, which is consistent with the observation in Zhao et al.¹⁸ indicating that the variability of coupling between different multiplicities might be less important than the adiabatic gap explored later. The T₁–S₀ SOC constants are shown to be relatively large (part of the results are given in Table 2 for the finally selected candidates). Since the T₁–S₀ coupling is responsible for the nonradiative decay of the triplet, we set the upper limit of T₁–S₀ SOC to be 15 cm⁻¹ to suppress the possible nonradiative decay so that the RISC can be promoted. The lower limit of the T₁–S₁ coupling is prescribed to be 0.1 cm⁻¹ so that a non-negligible intersystem SOC is guaranteed. Such screening further leads to 47 candidates for which the excited-state optimizations are performed.

The geometry optimization at the excited state is conducted at the level of TDDFT/B3LYP/def2-SVP. We notice that the selection of density functionals may give rise to a significant diversity of the computed excitation energies. The employment of usual exchange–correlation (XC) functionals with lower Hartree–Fock (HF) fraction, such as PBE0 and B3LYP, leads to an underestimation of the excitation energies,⁴² while functionals with a higher HF fraction, such as M062X, overestimate the excitation energies and give an enlarged singlet–triplet gap.⁴³ Experimental identification of the excitation energy is often done via the spectroscopic measurement, but the situation of the experiment–theory matching depends on the individual molecule. Furthermore,

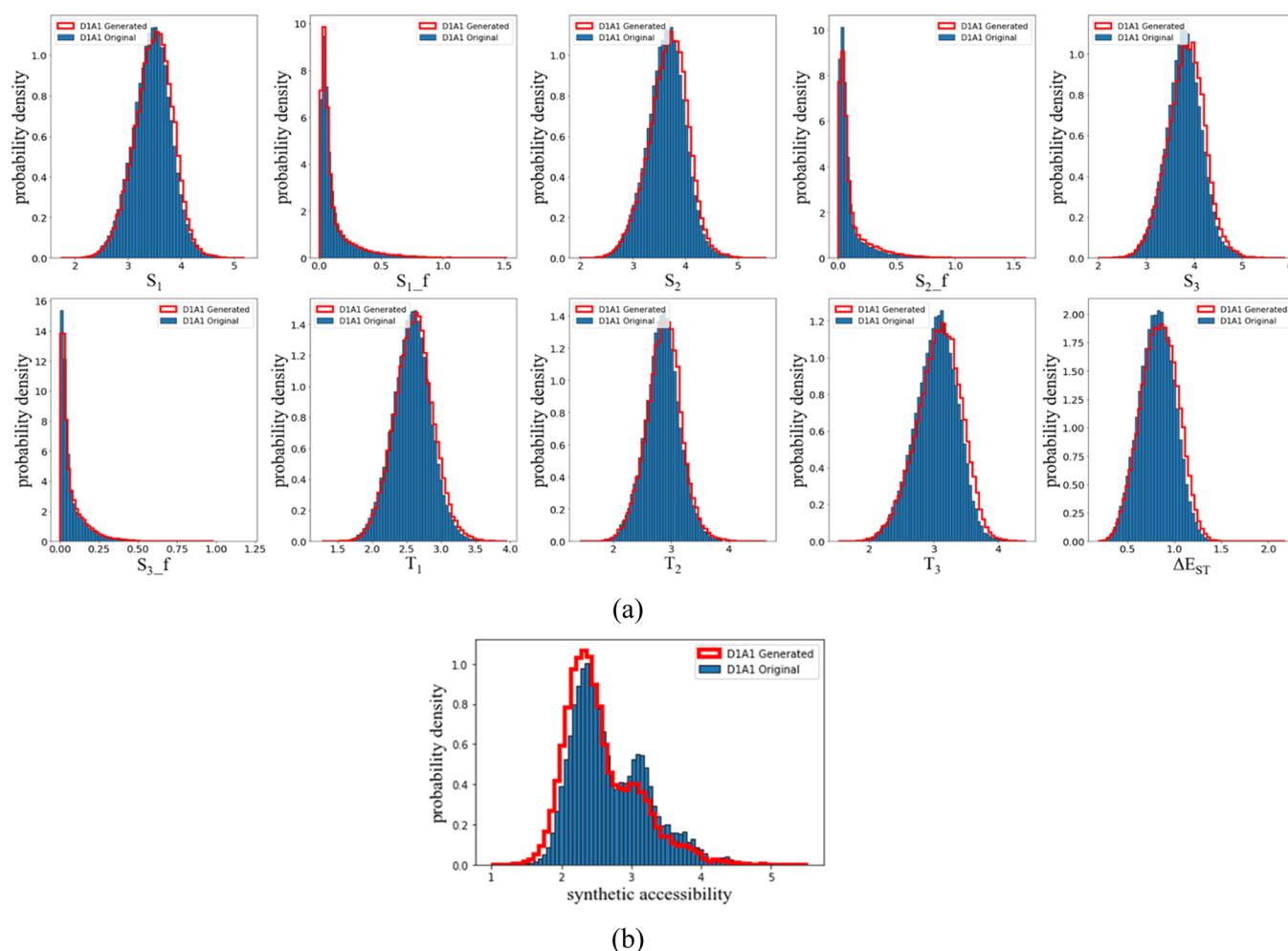


Figure 3. (a) Distributions of the excited-state properties of the original and generated D1A1 molecules. Note that both the original and generated properties are computed via the DNN model predictions. The singlet and triplet energies and the energy difference are in the unit of eV, while f is dimensionless. (b) Distributions of synthetic accessibility of the original and generated D1A1 molecules.

the empirical ΔE_{ST} values are known to have significant experimental uncertainty.¹⁷ We hereby claim that we only deal with the B3LYP-derived excited-state properties in this article, by particularly focusing on the reproducibility of the generative algorithm for molecular electronic structures.

S_1 and T_1 geometry optimizations for the 47 candidates are implemented in the TDDFT context. We also test the effect of TDA since it is believed to produce a more accurate triplet state.³⁶ The adiabatic singlet–triplet gaps with and without TDA are computed, with the scattered plot being shown in Figure S3 (in Supporting Information). It is observed that the B3LYP-TDA gap is mostly smaller than the B3LYP gap, demonstrating that the screening based on the B3LYP gap would be reasonable as a more accurate description of excited states can lead to even smaller gaps, which further justifies the selection criterion. Table 2 exhibits the samples with B3LYP gaps smaller than 0.25 eV, where the TDA gaps, T_1S_1 and T_1S_0 SOC constants, are displayed as well. These 19 molecules can be considered as the qualified candidates that are eventually selected from the generation space.

To validate that the delayed fluorescence originates from the D–A molecular design in the generation space, we carefully examine the molecular structure for the 19 TADF candidates, as shown in Figure 4. For some samples, either acceptor or donor fragment already exists in the empirical donor/acceptor

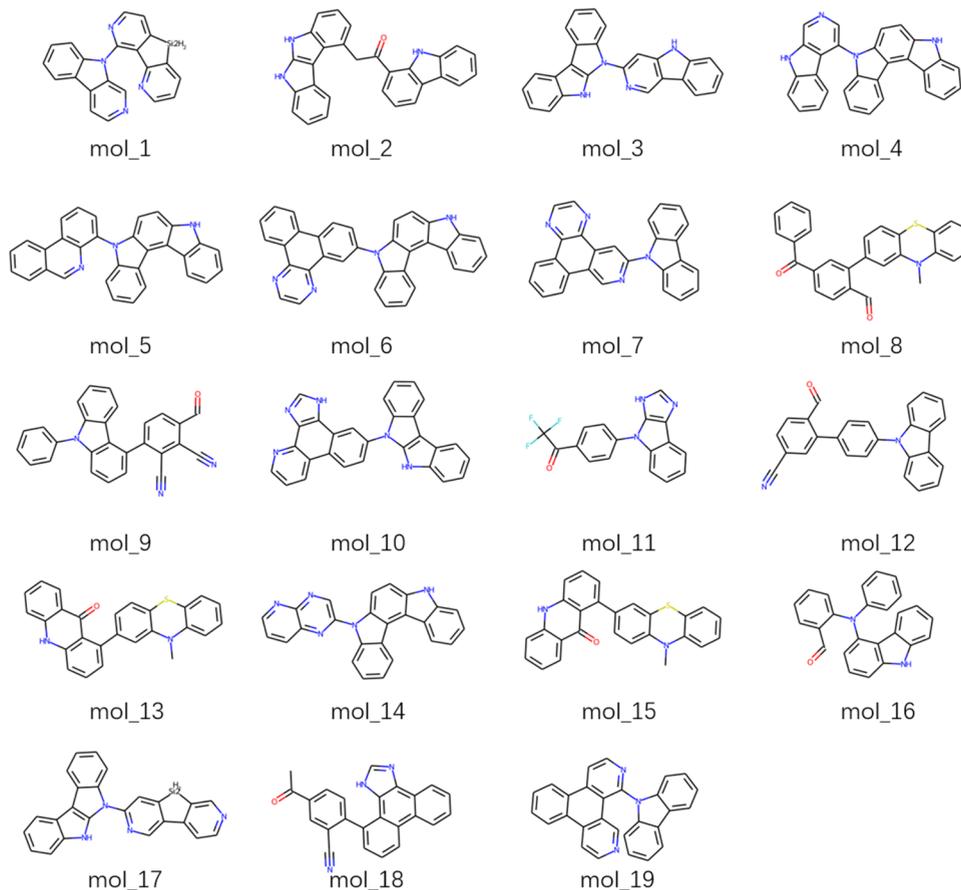
lists (Tables S1 and S2); however, the molecule holds the newly created donor or acceptor respectively. For example, in mol_8, a phenothiazine-like fragment serves as a known donor, which is connected to a novel acceptor. In another case, both fragments appear in the empirical list but simultaneously belong to the donor family or the acceptor family, such as mol_6, mol_13, and mol_15. The results indicate that the existing donor/acceptor lists are extensible based on the adversarial generated molecules.

It should be noted that the designed TADF candidates are of essentially D–A structure, since these molecules are screened out of the generation space, which is structurally similar to the synthesized original space, as a result of the adversarial generation algorithm. This actually imposes some limitations to the design strategy that we can only find molecular types that are predefined within the original space. For the current design case, multiple potential TADF configurations, e.g., multiresonance and so on, are therefore missing. The issue can be possibly resolved by more advanced algorithms where the transfer of chemical space can happen during the model optimization, which is subject to future research.

The frontier orbital analysis (calculated at the level of B3LYP/def2-SVP) is also performed for the 19 screened candidates at the optimized S_0 , S_1 , and T_1 geometries, with two typical examples shown in Figure 5. It is clearly seen that the

Table 2. Adiabatic Singlet–Triplet Energy Gaps Computed via the B3LYP Functional with and without TDA and the T_1S_1 and T_1S_0 SOC Constants at the S_0 Geometry for the Eventually Screened TADF Candidates

name	SMILES	B3LYP gap (eV)	B3LYP-TDA gap (eV)	T_1S_1 SOC (cm^{-1})	T_1S_0 SOC (cm^{-1})
mol_1	<chem>c1cnc2c(c1)[SiH2:2]c1cnc(-n3c4cccc4c4cnc43)c1-2</chem>	0.1333	0.0106	0.3239	1.8620
mol_2	<chem>O=C(Cc1cccc2[nH]c3[nH]c4cccc4c3c12)c1cccc2c1[nH]c1cccc12</chem>	0.0163	0.0117	1.8729	7.1232
mol_3	<chem>c1ccc2c(c1)[nH]c1c2e2cccc2n1-c1cc2[nH]c3cccc3c2cn1</chem>	0.1223	0.0980	0.1136	0.7360
mol_4	<chem>c1ccc2c(c1)[nH]c1ccc3c(c4cccc4n3-c3ncc4[nH]c5cccc5c34)c12</chem>	0.1759	0.0798	0.1606	0.1746
mol_5	<chem>c1cc2c(cc1)[nH]c1ccc3c(c4cccc4n3-c3cccc4c3ncc3cccc34)c12</chem>	0.0227	0.0144	0.1703	1.3593
mol_6	<chem>c1ccc2c(c1)[nH]c1ccc3c(c4cccc4n3-c3ccc4c5cccc5c5nccnc5c4c3)c12</chem>	0.0130	0.0117	0.1225	1.0532
mol_7	<chem>c1ccc2c(c1)-c1cnc(-n3c4cccc4c4cccc43)cc1c1nccnc21</chem>	0.0995	0.0926	0.1396	1.6184
mol_8	<chem>O=C(c1ccc(C=O)c(-c2ccc3c(c2)Nc2cccc2S3)c1)c1cccc1</chem>	0.0483	0.0391	0.1507	8.6579
mol_9	<chem>O=Cc1ccc(-c2cccc3c2c2cccc2n3-c2cccc2)c(C#N)c1C#N</chem>	0.1280	0.0367	0.2102	0.7563
mol_10	<chem>c1cnc2c(c1)-c1ccc(-n3c4cccc4c4c5cccc5[nH]c43)cc1c1[nH]cnc21</chem>	0.0082	0.0078	0.2818	0.8635
mol_11	<chem>O=C(c1ccc(-n2c3cccc3c3nc[nH]c23)cc1)C(F)(F)F</chem>	0.0070	0.0065	0.1691	1.7473
mol_12	<chem>N#Cc1ccc(C=O)c(-c2ccc(-n3c4cccc4c4cccc43)cc2)c1</chem>	0.0084	-0.0066	0.1288	9.6389
mol_13	<chem>CN1c2cccc2Sc2ccc(-c3cccc4Nc5cccc5c(=O)c34)cc21</chem>	0.2109	0.1751	0.5778	6.8733
mol_14	<chem>c1ccc2c(c1)[nH]c1ccc3c(c4cccc4n3-c3ncc4ncccc4n3)c12</chem>	0.0239	0.0204	0.1868	1.7842
mol_15	<chem>CN1c2cccc2Sc2cc(-c3cccc4[nH]c5cccc5c(=O)c34)ccc21</chem>	0.1956	0.1594	0.8683	14.9823
mol_16	<chem>O=Cc1cccc1N(c1ccc2[nH]c3cccc3c12)c1cccc1</chem>	0.1579	0.1168	0.4755	6.4195
mol_17	<chem>c1ccc2c(c1)[nH]c1c2c2cccc2n1-c1ncc2c(c1)[SiH2:2]c1ncccc1-2</chem>	0.0761	0.0616	0.1962	0.7084
mol_18	<chem>CC(=O)c1ccc(-c2cccc3c4cccc4c4nc[nH]c4c23)c(C#N)c1</chem>	0.2286	0.1081	0.1208	1.2150
mol_19	<chem>c1cc2c(cn1)-c1c(ccnc1-n1c3cccc3c3cccc31)c1cccc12</chem>	0.2335	0.0899	0.5408	1.8974

**Figure 4.** Generated molecules that fulfill the fine screening criteria with satisfactory quantum chemical properties for the delayed fluorescence.

HOMO and LUMO orbitals are efficiently separated in both the ground and excited states, which is consistent with the minimal adiabatic gaps computed for these TADF molecules.

The S_1 and T_1 geometries are similar to each other, but the donor–acceptor dihedral twisting is notably enlarged (the case is more significant in mol_11). The observation is in line with

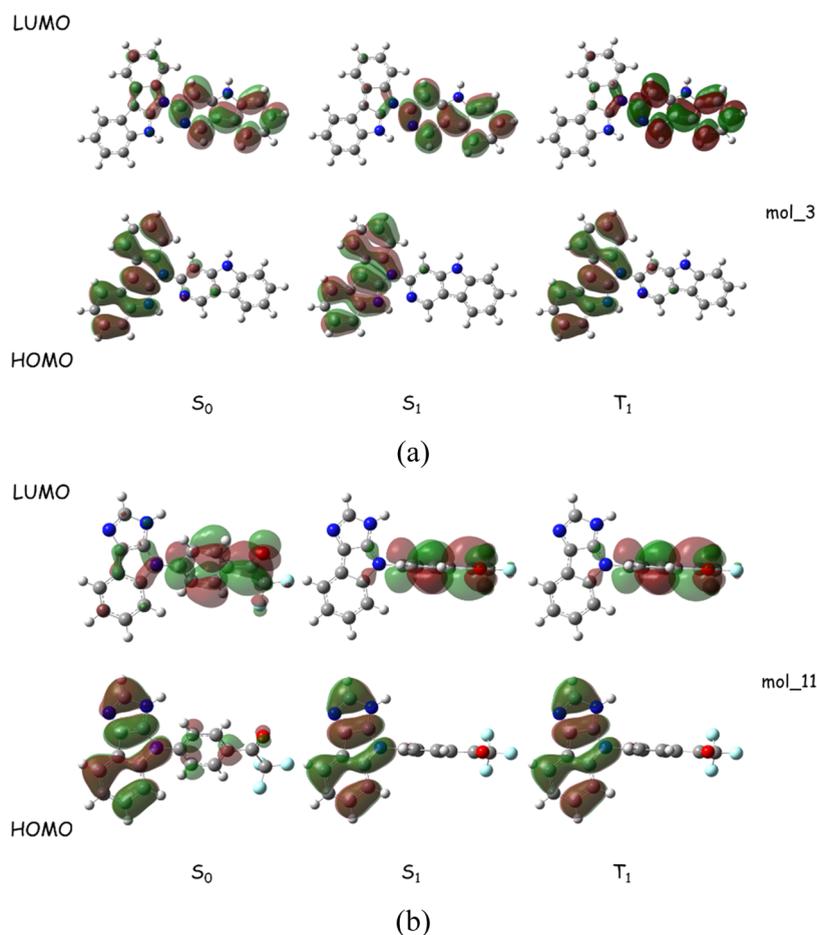


Figure 5. Frontier orbitals for two of the fine screened molecules analyzed at the optimized S_0 , S_1 , and T_1 geometries.

the principal TADF designing rule^{44–46} where a twisted D–A structure is built to realize the reduced orbital overlap and the minimized ΔE_{ST} .

CONCLUSIONS

The article has proposed an alternative design route for the TADF molecules beyond the current D–A combination and high-throughput screening. The integrated modules combining the algorithmic synthesis, deep prediction, adversarial generation, and fine screening, essentially provide a comprehensive paradigm for synthesizing and generating molecules with expected excited-state properties. The simple donor–acceptor design is adopted where the original D1A1 space is synthesized according to the BRICS rule. Nearly 14,000 samples are labeled via high-throughput calculations for the purpose of DNN model training and property prediction. Satisfactory forecasting precisions are obtained with deep learning, especially for the most important factors, the oscillator strength and the singlet–triplet splitting that to a large extent prescribe the TADF process. Adversarial autoencoder is applied to generate novel structures in terms of the molecular SMILES that can mimic the original D1A1 samples. With the assistance of DNN predictions, the distributions of excited-state properties for the generated molecules are found to perfectly match those of the original samples. Further analysis shows that the synthetic accessibility for both the original and generated D1A1 molecules is at a reasonable level, which facilitates material synthesis and device fabrication. By performing the

fine screening on the generated chemical space, a bundle of qualified TADF candidates is eventually selected with minimal adiabatic gaps and moderate spin–orbital couplings. In addition to the similar structures, the adversarial generation algorithm can well reproduce the charge-transfer characteristics and the donor–acceptor dihedral twisting at the excited state, indicating a strong potential of the generative model in designing materials with expected photophysical mechanisms.

Our research opens another pathway to the TADF molecular designs. Given the generative model and the original molecular space, the methodology can be easily applied to generate other types of TADF molecules, e.g., molecules with more complex D–A structures, molecules with resonant structures,⁴⁷ and molecules with the negative singlet–triplet gap⁴⁸ as recently introduced. The model can be further improved by incorporating reinforcement learning,^{39,49,50} with which a biased property distribution can be generated and moved along the preferred direction. We expect that the current framework can lead to a significant enrichment of the TADF library and a higher vision for “design as we desire” that is possibly realizable in the near future.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsomega.2c02253>.

Empirical acceptor and donor moieties list, DNN prediction qualities on the D1A1 data set, methodo-

logical details for quantum chemistry and machine learning, learning process of adversarial autoencoder, structural similarity analysis, internal diversity analysis, and adiabatic gaps with and without TDA (PDF)

AUTHOR INFORMATION

Corresponding Author

Zheng Tan – Chengdu Polytechnic, Chengdu, Sichuan 610000, P. R. China; orcid.org/0000-0001-9491-6770; Email: zhengtan1983@hotmail.com

Authors

Yan Li – Xiyuan Quantitative Technology, Chengdu, Sichuan 610000, P. R. China

Ziying Zhang – Guangzhou Yinfo Information Technology, Guangzhou 511431, P. R. China

Xin Wu – Xiyuan Quantitative Technology, Chengdu, Sichuan 610000, P. R. China

Thomas Penfold – Chemistry-School of Natural and Environmental Sciences, Newcastle University, Newcastle Upon Tyne NE1 7RU, U.K.

Weimei Shi – Chengdu Polytechnic, Chengdu, Sichuan 610000, P. R. China

Shiqing Yang – Chengdu Polytechnic, Chengdu, Sichuan 610000, P. R. China

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsomega.2c02253>

Notes

The authors declare no competing financial interest. All of the codes associated with the research work are made available at <https://github.com/XiyuanQuantitative-Technology/Adversarial-generation>. Relevant methodological details during the implementation are given in Supporting Information.

ACKNOWLEDGMENTS

The work is supported by the Sichuan Science and Technology Project (No. 2019YJ0646), the Chengdu Science and Technology Project (No. 2019-YF05-00224-SN), the Research Platform Foundation of Chengdu Polytechnic (Nos. 19KYPT01 and 20KYTD07).

REFERENCES

- (1) Uoyama, H.; Goushi, K.; Shizu, K.; Nomura, H.; Adachi, C. Highly efficient organic light-emitting diodes from delayed fluorescence. *Nature* **2012**, *492*, 234–238.
- (2) Wong, M. Y.; Zysman-Colman, E. Purely Organic Thermally Activated Delayed Fluorescence Materials for Organic Light-Emitting Diodes. *Adv. Mater.* **2017**, *29*, No. 1605444.
- (3) Ravinson, D. S. M.; Thompson, M. E. Thermally assisted delayed fluorescence (TADF): fluorescence delayed is fluorescence denied. *Mater. Horiz.* **2020**, *7*, 1210–1217.
- (4) Huang, S. P.; Zhang, Q. S.; Shiota, Y.; Nakagawa, T.; Kuwabara, K.; Yoshizawa, K.; Adachi, C. Computational prediction for singlet- and triplet-transition energies of charge-transfer compounds. *J. Chem. Theory Comput.* **2013**, *9*, 3872–3877.
- (5) Yang, Z. Y.; Mao, Z.; Xie, Z. L.; Zhang, Y.; Liu, S. W.; Zhao, J.; Xu, J. R.; Chi, Z. G.; Aldred, M. P. Recent advances in organic thermally activated delayed fluorescence materials. *Chem. Soc. Rev.* **2017**, *46*, 915–1016.
- (6) Liang, X.; Tu, Z. L.; Zheng, Y. X. Thermally Activated Delayed Fluorescence Materials: Towards Realization of High Efficiency

Through Strategic Small Molecular Design. *Chem. - Eur. J.* **2019**, *25*, 5623–5642.

(7) Kim, J. H.; Yun, J. H.; Lee, J. Y. Recent Progress of Highly Efficient Red and Near-Infrared Thermally Activated Delayed Fluorescent Emitters. *Adv. Opt. Mater.* **2018**, *6*, No. 1800255.

(8) Huang, T. Y.; Jiang, W.; Duan, L. Recent progress in solution processable TADF materials for organic light-emitting diodes. *J. Mater. Chem. C* **2018**, *6*, 5577–5596.

(9) Penfold, T. J.; Dias, F. B.; Monkman, A. P. The theory of thermally activated delayed fluorescence for organic light emitting diodes. *Chem. Commun.* **2018**, *54*, 3926–3935.

(10) Sun, L. J.; Hua, W. J.; Liu, Y.; Tian, G. J.; Chen, M. X.; Chen, M. X.; Yang, F. X.; Wang, S. F.; Zhang, X. T.; Luo, Y.; Hu, W. P. Thermally Activated Delayed Fluorescence in an Organic Cocystal: Narrowing the Singlet-Triplet Energy Gap via Charge Transfer. *Angew. Chem., Int. Ed.* **2019**, *58*, 11311–11316.

(11) Wang, H.; Xie, L. S.; Peng, Q.; Meng, L. Q.; Wang, Y.; Yi, Y. P.; Wang, P. F. Novel Thermally Activated Delayed Fluorescence Materials—Thioxanthone Derivatives and Their Applications for Highly Efficient OLEDs. *Adv. Mater.* **2014**, *26*, 5198–5204.

(12) Tanaka, H.; Shizu, K.; Lee, J. Y.; Adachi, C. Effect of Atom Substitution in Chalcogenodiazole-Containing Thermally Activated Delayed Fluorescence Emitters on Radiationless Transition. *J. Phys. Chem. C* **2015**, *119*, 2948–2955.

(13) Lee, J. Y.; Shizu, K.; Tanaka, H.; et al. Controlled emission colors and singlet-triplet energy gaps of dihydrophenazine-based thermally activated delayed fluorescence emitters. *J. Mater. Chem. C* **2015**, *3*, 2175–2181.

(14) Lee, D. R.; Hwang, S. H.; Jeon, S. K.; Lee, C. W.; Lee, J. Y. Benzofurocarbazole and benzothienocarbazole as donors for improved quantum efficiency in blue thermally activated delayed fluorescent devices. *Chem. Commun.* **2015**, *51*, 8105–8107.

(15) Shu, Y.; Levine, B. G. Simulated evolution of fluorophores for light emitting diodes. *J. Chem. Phys.* **2015**, *142*, No. 104104.

(16) Zhang, X.; Shen, W.; Zhang, D.; Zheng, Y.; He, R.; Li, M. Theoretical investigation of dihydroacridine and diphenylsulphone derivatives as thermally activated delayed fluorescence emitters for organic light-emitting diodes. *RSC Adv.* **2015**, *5*, 51586–51591.

(17) Gómez-Bombarelli, R.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Duvenaud, D.; Maclaurin, D.; Blood-Forsythe, M. A.; Chae, H. S.; Einzinger, M.; Ha, D. G.; Wu, T.; Markopoulos, G.; et al. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nat. Mater.* **2016**, *15*, 1120–1127.

(18) Zhao, K.; Omar, O. H.; Nemataram, T.; Padula, D.; Troisi, A. Novel thermally activated delayed fluorescence materials by high-throughput virtual screening: going beyond donor-acceptor design. *J. Mater. Chem. C* **2021**, *9*, 3324.

(19) Méndez-Lucio, O.; Baillif, B.; Clevert, D. A.; Rouquie, D.; Wichard, J. De novo generation of hit-like molecules from gene expression signatures using artificial intelligence. *Nat. Commun.* **2020**, *11*, No. 10.

(20) Prykhodko, O.; Johansson, V. S.; Kotsias, P. C.; Arús-Pous, J.; Bjerrum, E. J.; Engkvist, O.; Chen, H. M. A de novo molecular generation method using latent vector based generative adversarial network. *J. Cheminf.* **2019**, *11*, No. 74.

(21) Blaschke, T.; Olivecrona, M.; Engkvist, O.; Bajorath, J.; Chen, H. M. Application of Generative Autoencoder in De Novo Molecular Design. *Mol. Inf.* **2018**, *37*, No. 1700123.

(22) Kadurin, A.; Nikolenko, S.; Khrabrov, K.; Aliper, A.; Zhavoronkov, A. druGAN: An Advanced Generative Adversarial Autoencoder Model for de Novo Generation of New Molecules with Desired Molecular Properties in Silico. *Mol. Pharmaceutics* **2017**, *14*, 3098–3104.

(23) Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **2018**, *361*, 360–365.

- (24) 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. 2018.
- (25) Sanchez-Lengeling, B.; Outeiral, C.; Guimaraes, G. L.; Aspuru-Guzik, A. Optimizing distributions over molecular space. An Objective-Reinforced Generative Adversarial Network for Inverse-design Chemistry (ORGANIC). 2017.
- (26) Gao, K. F.; Nguyen, D. D.; Tu, M. H.; Wei, G. W. Generative Network Complex for the Automated Generation of Drug-like Molecules. *J. Chem. Inf. Model* **2020**, *60*, 5682–5698.
- (27) Wegscheid-Gerlach, C.; Zaliani, A.; et al. On the Art of Compiling and Using 'Drug-Like' Chemical Fragment Spaces. *ChemMedChem* **2008**, *3*, 1503–1507.
- (28) Tosco, P.; Stiefl, N.; Landrum, G. Bringing the MMFF force field to the RDKit: implementation and validation. *J. Cheminf.* **2014**, *6*, No. 37.
- (29) 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*, Revision A.03; Gaussian, Inc.: Wallingford CT, 2016.
- (30) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model* **2010**, *50*, 742–754.
- (31) 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. *Front. Pharmacol.* **2020**, *11*, No. 565644.
- (32) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36.
- (33) Williams, R. J.; Zipser, D. A Learning Algorithm for Continually Running Fully Recurrent Neural Networks. *Neural Comput.* **1989**, *1*, 270–280.
- (34) Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA quantum chemistry program package. *J. Chem. Phys.* **2020**, *152*, No. 224108.
- (35) Hirata, S.; Head-Gordon, M. Time-dependent density functional theory within the Tamm-Dancoff approximation. *Chem. Phys. Lett.* **1999**, *314*, 291–299.
- (36) Peach, M. J. G.; Williamson, M. J.; Tozer, D. J. Influence of Triplet Instabilities in TDDFT. *J. Chem. Theory Comput.* **2011**, *7*, 3578–3585.
- (37) Lu, C. Q.; Liu, Q.; Sun, Q.; Hsieh, C. Y.; Zhang, S. Y.; Shi, L.; Lee, C. K. Deep Learning for Optoelectronic Properties of Organic Semiconductors. *J. Phys. Chem. C* **2020**, *124*, 7048–7060.
- (38) Ertl, P.; Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. Cheminf.* **2009**, *1*, No. 8.
- (39) Popova, M.; Isayev, O.; Tropsha, A. Deep reinforcement learning for de novo drug design. *Sci. Adv.* **2018**, *4*, No. eaap7885.
- (40) Kim, K.; Kang, S.; Yoo, J.; Kwon, Y. C.; Nam, Y.; Lee, D.; Kim, I.; Choi, Y. S.; Jung, Y.; Kim, S.; et al. Deep-learning-based inverse design model for intelligent discovery of organic molecules. *npj Comput. Mater.* **2018**, *4*, No. 67.
- (41) Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; S 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.
- (42) Penfold, T. J.; Dias, F. B.; Monkman, A. P. The theory of thermally activated delayed fluorescence for organic light emitting diodes. *Chem. Commun.* **2018**, *54*, 3926–3935.
- (43) Peng, Q.; Fan, D.; Duan, R. H.; Yi, Y. P.; Niu, Y. L.; Wang, D.; Shuai, Z. G. Theoretical Study of Conversion and Decay Processes of Excited Triplet and Singlet States in a Thermally Activated Delayed Fluorescence Molecule. *J. Phys. Chem. C* **2017**, *121*, 13448–13456.
- (44) Zhang, Y. W.; Ma, H. L.; Wang, S. P.; Li, Z. Q.; Ye, K. Q.; Zhang, J. Y.; Liu, Y.; Peng, Q.; Wang, Y. Supramolecular Structure-Dependent Thermally-Activated Delayed Fluorescence (TADF) Properties of Organic Polymorphs. *J. Phys. Chem. C* **2016**, *120*, 19759–19767.
- (45) Tanaka, H.; Shizu, K.; Nakanotani, H.; Adachi, C. Twisted Intramolecular Charge Transfer State for Long-Wavelength Thermally Activated Delayed Fluorescence. *Chem. Mater.* **2013**, *25*, 3766–3771.
- (46) Wu, K. L.; Zhang, T.; Zhan, L.; et al. Tailoring Optoelectronic Properties of Phenanthroline-Based Thermally Activated Delayed Fluorescence Emitters through Isomer Engineering. *Adv. Opt. Mater.* **2016**, *4*, 1558–1566.
- (47) Eng, J.; Penfold, T. J. Open questions on the photophysics of thermally activated delayed fluorescence. *Commun. Chem.* **2021**, *4*, No. 91.
- (48) Pollice, R.; Friederich, P.; Lavigne, C.; Gomes, G. D. P.; Aspuru-Guzik, A. Organic molecules with inverted gaps between first excited singlet and triplet states and appreciable fluorescence rates. *Matter* **2021**, *4*, 1654–1682.
- (49) Pereira, T.; Abbasi, M.; Ribeiro, B.; Arrais, J. P. Diversity oriented Deep Reinforcement Learning for targeted molecule generation. *J. Cheminf.* **2021**, *13*, No. 21.
- (50) Zhou, Z. P.; Kearnes, S.; Li, L.; Zare, R. N.; Riley, P. Optimization of Molecules via Deep Reinforcement Learning. *Sci. Rep.* **2019**, *9*, No. 10752.