

# Deep Learning Optical Spectroscopy Based on Experimental Database: Potential Applications to Molecular Design

Joonyoung F. Joung,<sup>†</sup> Minhi Han,<sup>†</sup> Jinhyo Hwang, Minseok Jeong, Dong Hoon Choi, and Sunnam Park<sup>\*</sup>

Cite This: *JACS Au* 2021, 1, 427–438

Read Online

ACCESS |

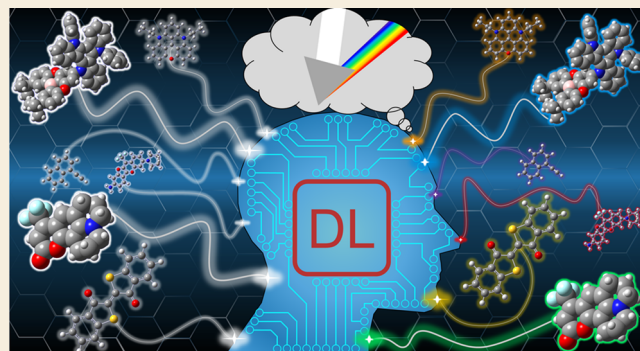
Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** Accurate and reliable prediction of the optical and photophysical properties of organic compounds is important in various research fields. Here, we developed deep learning (DL) optical spectroscopy using a DL model and experimental database to predict seven optical and photophysical properties of organic compounds, namely, the absorption peak position and bandwidth, extinction coefficient, emission peak position and bandwidth, photoluminescence quantum yield (PLQY), and emission lifetime. Our DL model included the chromophore–solvent interaction to account for the effect of local environments on the optical and photophysical properties of organic compounds and was trained using an experimental database of 30 094 chromophore/solvent combinations. Our DL optical spectroscopy made it possible to reliably and quickly predict the aforementioned properties of organic compounds in solution, gas phase, film, and powder with the root mean squared errors of 26.6 and 28.0 nm for absorption and emission peak positions, 603 and 532 cm<sup>-1</sup> for absorption and emission bandwidths, and 0.209, 0.371, and 0.262 for the logarithm of the extinction coefficient, PLQY, and emission lifetime, respectively. Finally, we demonstrated how a blue emitter with desired optical and photophysical properties could be efficiently virtually screened and developed by DL optical spectroscopy. DL optical spectroscopy can be efficiently used for developing chromophores and fluorophores in various research areas.

**KEYWORDS:** deep learning, optical property, photophysical property, fluorophore, chromophore, chromophore-solvent interaction, experimental database



## INTRODUCTION

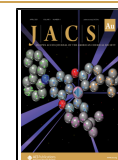
The optical and photophysical properties of chromophores and fluorophores are important for industrial and academic applications in a variety of research fields such as, organic solar cells,<sup>1</sup> organic light-emitting diodes (OLEDs),<sup>2</sup> bioimaging,<sup>3</sup> organic sensors,<sup>4</sup> and organic dyes. Substantial efforts have been made to design and synthesize new colorful materials based on various chromophores and fluorophores for various purposes. To develop new molecules with desired properties, knowledge of the chemical and physical properties of previously studied molecules is required. In particular, the optical properties of molecules, such as absorption and emission peak positions and bandwidths, extinction coefficients, photoluminescence quantum yield (PLQY), and emission lifetime, are important for the development of chromophores and fluorophores. For given molecules, the aforementioned optical properties are readily measured using various spectroscopic instruments, such as spectrophotometers, spectrofluorometers with integrating spheres, and time-resolved fluorescence (TRF) spectrometers. It is also desirable that the optical properties of newly designed chromophores and fluorophores are accurately

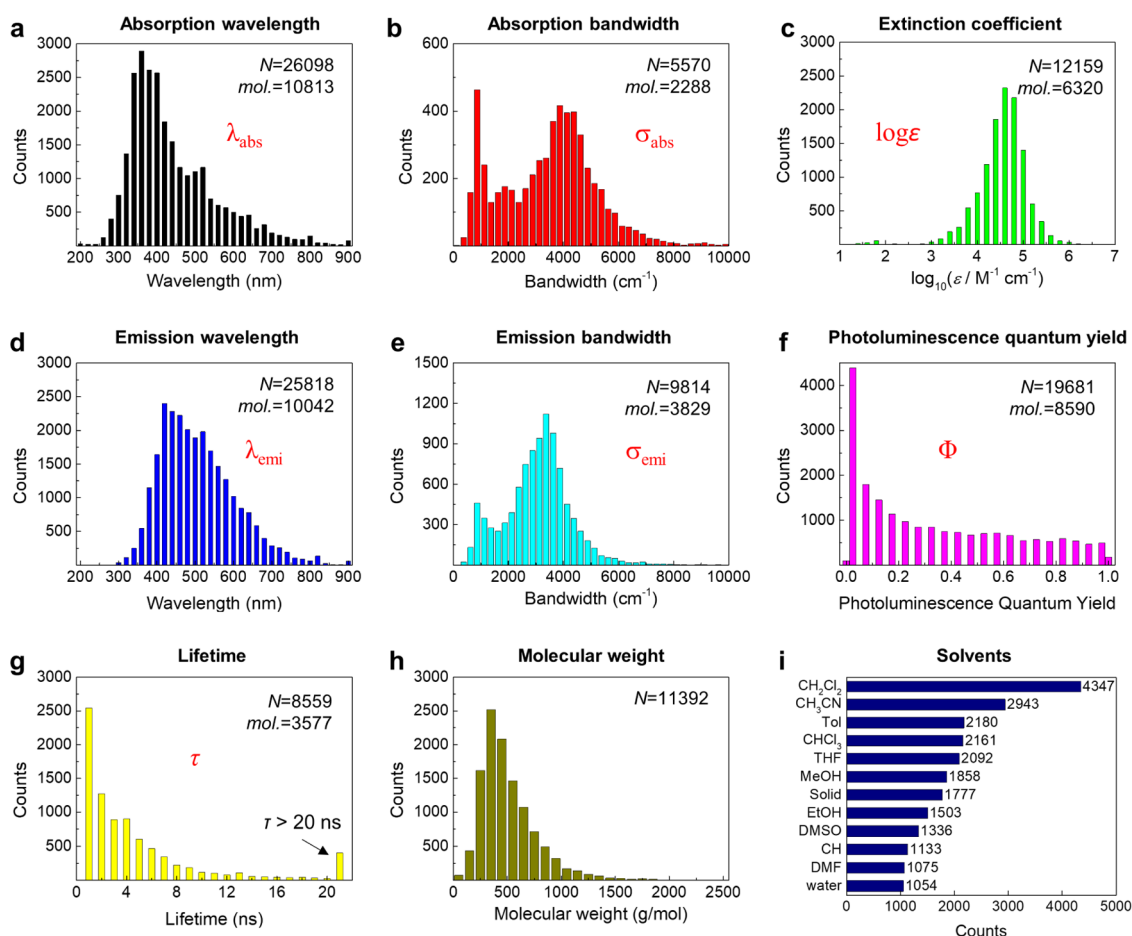
estimated in advance to reduce development costs. To estimate the optical properties of molecules, empirical rules or theories, such as the Woodward–Fieser rule,<sup>5</sup> first-principles,<sup>6,7</sup> and time-dependent density functional theory (TD-DFT) calculations<sup>8–10</sup> have been developed. Such methods have limitations resulting from the limited range of molecules to which they can be applied and from the lack of practical theories to calculate specific properties. While these rules and theories have been used to successfully predict many chemical and physical properties, more accurate, facile, and computationally inexpensive methods are still needed.

In this study, we have developed a deep learning (DL) method to reliably and quickly predict the optical and photophysical properties of organic compounds. Our DL

Received: January 29, 2021

Published: March 17, 2021





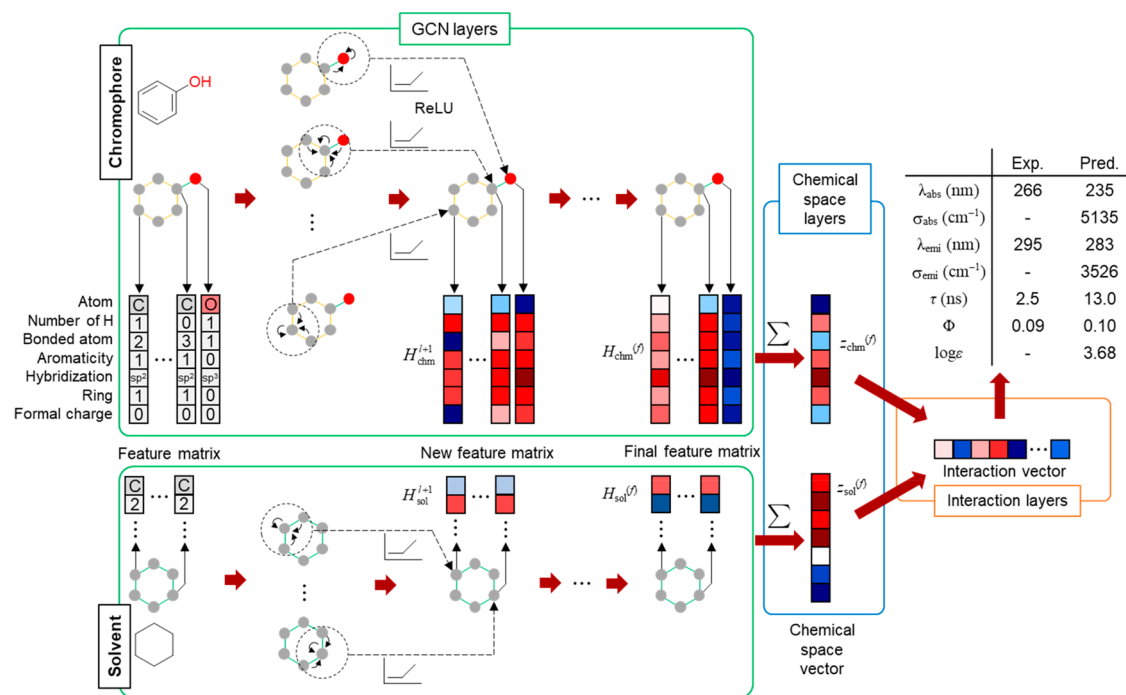
**Figure 1.** Database of the optical properties of organic compounds. Histograms of (a) first absorption peak position ( $\lambda_{\text{abs}}$ ), (b) bandwidth in full width at half-maximum ( $\sigma_{\text{abs}}$ ), (c) extinction coefficient in logarithm ( $\log \epsilon$ ), (d) emission peak position ( $\lambda_{\text{emi}}$ ), (e) bandwidth in full width at half-maximum ( $\sigma_{\text{emi}}$ ), (f) photoluminescence quantum yield ( $\Phi$ ), (g) lifetime ( $\tau$ ), (h) molecular weights of chromophores, and (i) solvents ( $\text{CH}_2\text{Cl}_2$  dichloromethane,  $\text{CH}_3\text{CN}$  acetonitrile, Tol toluene,  $\text{CHCl}_3$  chloroform, THF tetrahydrofuran, MeOH methanol, EtOH ethanol, DMSO dimethyl sulfoxide, CH cyclohexane, and DMF *N,N*-dimethylformamide). The number of data points ( $N$ ) and the number of chromophores ( $\text{mol.}$ ) are included in each graph.<sup>11,12</sup>

method includes chromophore–solvent interactions to account for the effect of environment on the optical and photophysical properties. Our DL model was trained using an experimental database of 30 094 chromophore/solvent combinations consisting of 11 392 unique organic chromophores in 369 solvents or solid states.<sup>11,12</sup> Our DL method is found to readily predict seven optical properties such as the first absorption peak position and bandwidth, extinction coefficient, emission peak position and bandwidth, photoluminescence quantum yield, and emission lifetime. The root mean squared errors of the predicted values were found to be 26.6 and 28.0 nm for absorption and emission peak positions, 603 and 532  $\text{cm}^{-1}$  for absorption and emission bandwidths, and 0.209, 0.371, and 0.262 for the logarithm of the extinction coefficient, PLQY, and emission lifetime, respectively. Using these predicted optical properties of organic compounds, the electronic absorption and emission spectra are shown to be readily reconstructed, and the colors of organic molecules associated with the absorption and emission are estimated in Commission Internationale de l’Eclairage (CIE) color coordinates. Finally, we demonstrate that the DL model can be effectively utilized to prescreen newly designed molecules to find target molecules that exhibit the desired optical and photophysical properties.

## RESULTS AND DISCUSSION

### Experimental Optical Properties of Organic Compounds

The absorption properties of a chromophore in a solvent are characterized by the first absorption peak position ( $\lambda_{\text{abs}}$ ), bandwidth ( $\sigma_{\text{abs}}$ ), and extinction coefficient ( $\epsilon$ ), whereas the emission properties of a fluorophore are characterized by the emission peak position ( $\lambda_{\text{emi}}$ ), bandwidth ( $\sigma_{\text{emi}}$ ), and photoluminescence quantum yield (PLQY,  $\Phi$ ). In addition, the emission (excited-state) lifetime ( $\tau$ ) is an intrinsic property of fluorophores that is importantly considered in fluorescence lifetime imaging and dynamic quenching. The experimental database of these optical and photophysical properties of organic compounds is prerequisite for our DL method. As shown in Figure 1, the experimental database includes 11 392 unique organic chromophores in 369 solvents or solid states, yielding 30 094 chromophore/solvent combinations. Briefly, the experimental database was built by collecting seven optical and photophysical properties of organic compounds from the literature and has been described in detail elsewhere.<sup>11,12</sup> Not all seven optical and photophysical properties were available for each chromophore/solvent pair in the database. A variety of chromophore structures are included in the database; the frequencies of specific core structures are summarized in Figure



**Figure 2.** Our deep learning (DL) model. The interaction vector is used to account for the chromophore–solvent interaction for predicting the optical and photophysical properties of the chromophore.

S1. For example, more than 1300 BF<sub>2</sub>-group-containing boron-dipyromethene (BODIPY) derivatives are found in the database.

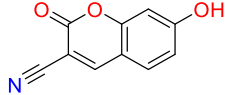
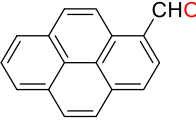
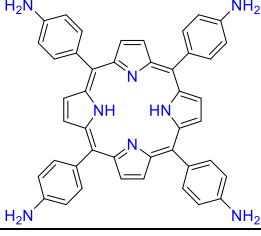
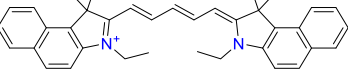
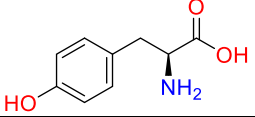
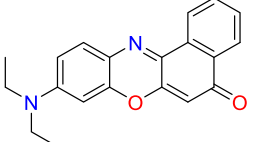
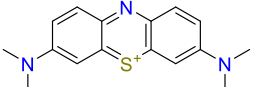
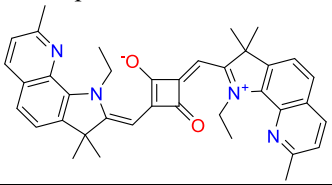
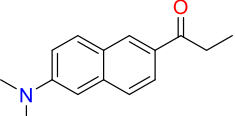
### DL Model for Prediction of Optical and Photophysical Properties

DL based on big data has attracted substantial attention because it can potentially solve problems via direct learning from data without well-defined rules, empirical laws, or theories.<sup>13,14</sup> In chemistry, DL has proven promising for predicting various properties of molecules and materials,<sup>15–29</sup> optimizing reactions and retrosynthesis,<sup>30–32</sup> and designing new molecules including drugs<sup>33,34</sup> and materials.<sup>35–37</sup> Additionally, DL has been effectively used in computational chemistry.<sup>38,39</sup> Nakata and Shimazaki reported a large-scale electronic structure database (PubChemQC)<sup>40</sup> and recently, using the PubChemQC database, Lee and co-workers reported a random forest model to predict the highest oscillator strength and the corresponding excitation energy of molecules.<sup>26</sup> Lin and co-workers used a DL method to predict the bandgap of configurationally hybridized graphene and boron nitride,<sup>21</sup> and Chang and co-workers used a tuplewise material representation to predict the band gap of organic–inorganic perovskite, 2D materials, and binary and ternary inorganic materials.<sup>27</sup> Jiang and co-workers reported DL models that could predict infrared and ultraviolet absorption spectra from the conformations of a molecule using a theoretical database that was generated by molecular dynamics simulations and DFT calculations.<sup>23,28</sup> Rinke and co-workers reported a DL method to predict molecular excitation spectra based on the QM7b and QM9 data sets of small organic molecules containing up to 17 C, N, O, S, and halogen atoms.<sup>24</sup> Li and co-workers reported a machine learning method using the Morgan fingerprint and an experimental database of 4300 molecules to predict three properties such as absorption and emission wavelengths and PLQY.<sup>29</sup>

In this work, we developed a DL model to predict the optical and photophysical properties as described in Figure 2. The molecular structure of a chromophore is represented by a  $150 \times 150$  adjacency matrix ( $A_{\text{chm}}$ ) and a  $150 \times 43$  feature matrix ( $X_{\text{chm}}$ ) (Figure S2). The adjacency matrix includes the connectivity information on atoms in the chromophore, and the feature matrix ( $X_{\text{chm}}$ ) includes the identity of the atoms, the number of hydrogen and heavy atoms bonded to them, aromaticity, hybridization, ring structure, and formal charge. The graph convolutional network (GCN) is the simplest version of the message-passing neural network, which is effective for graph structures such as molecular structures.<sup>16,25,41–44</sup> A single graph convolution of the adjacency matrix ( $A_{\text{chm}}$ ) and feature matrix of the  $l$ th layer ( $H_{\text{chm}}^l$ ) updates the atom's features, producing a new feature matrix for the  $l + 1$ th layer,  $H_{\text{chm}}^{l+1}$ , which contains all the information related to the nearest-neighboring atoms in the chromophore. The final feature matrix of GCN layers,  $H_{\text{chm}}^{(l)}$ , is reduced to a row vector,  $z_{\text{chm}}^{(l)}$ , by summing all elements to secure permutation invariance. The row vector,  $z_{\text{chm}}^{(l)}$ , passes through multilayer perceptrons (MLPs) to produce a chemical space vector,  $z_{\text{chm}}^{(f)}$ , for the chromophore. Similarly, a chemical space vector,  $z_{\text{sol}}^{(f)}$ , of the solvent is obtained. The chemical space vectors for the chromophore and solvent,  $z_{\text{chm}}^{(f)}$  and  $z_{\text{sol}}^{(f)}$ , are concatenated and pass through the MLPs to construct an interaction vector. Finally, the interaction vector passes through the MLPs to obtain the seven optical properties. Note that the interaction vector in the DL model is used to account for the chromophore–solvent interaction.

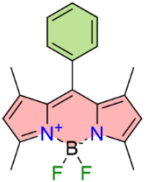
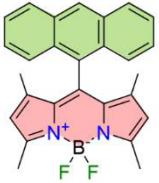
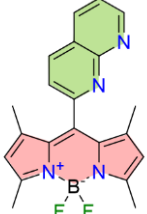
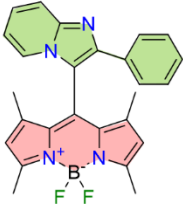
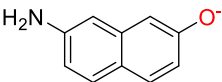
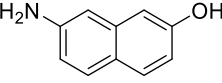
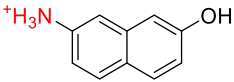
It should be highlighted that our DL model includes the chromophore–solvent interaction for accurate estimation of the optical and photophysical properties of organic compounds in various solvents. Details of the DL model and training process are provided in the Supporting Information and Methods. The performance of our DL model is presented in Figure S4. Overall, the DL model is reasonably well trained using the experimental database and the square of the Pearson correlation coefficient

Table 1. Prediction of Optical Properties Using Our DL Model<sup>a</sup>

Molecular structures		$\lambda_{\text{abs}}$ (nm)	$\sigma_{\text{abs}}$ ( $\text{cm}^{-1}$ )	$\log \epsilon$	$\lambda_{\text{emi}}$ (nm)	$\sigma_{\text{emi}}$ ( $\text{cm}^{-1}$ )	$\Phi$	$\tau$ (ns)
Coumarin in water 	Exp.	355	-	-	410	-	-	2.80
	Pred.	364	4854	4.27	442	3857	0.013	1.10
Pyrene in DMF 	Exp.	390	-	-	418	2966	0.002	64.8
	Pred.	388	4050	4.37	446	3285	0.007	7.7
Porphyrin in DMF 	Exp.	670	-	3.93	686	961	0.18	7.49
	Pred.	567	2360	4.57	624	1582	0.20	2.3
Cyanine in $\text{CH}_2\text{Cl}_2$ 	Exp.	396	875	-	710	735	-	1.3
	Pred.	688	1032	5.19	713	816	0.22	1.2
L-tyrosine in water 	Exp.	285	-	-	300	-	-	3.27
	Pred.	276	4865	2.74	328	4640	0.023	2.7
Nile Red in methanol 	Exp.	549	2960	-	638	1783	0.62	3.65
	Pred.	548	2473	4.66	615	2105	0.38	2.7
Methylene blue in water 	Exp.	664	1817	4.59	683	886	0.02	3.54
	Pred.	672	1023	4.93	698	879	0.06	0.31
Squaraine in $\text{CHCl}_3$ 	Exp.	678	1046	5.19	702	818	0.6	-
	Pred.	662	1201	5.11	692	1006	0.225	1.5
Prodan in hexane 	Exp.	338	-	-	389	3142	0.02	0.29
	Pred.	349	4401	4.53	398	3216	0.08	1.1

a

Table 1. continued

	Molecular structures		$\lambda_{\text{abs}}$ (nm)	$\sigma_{\text{abs}}$ ( $\text{cm}^{-1}$ )	$\log \epsilon$	$\lambda_{\text{emi}}$ (nm)	$\sigma_{\text{emi}}$ ( $\text{cm}^{-1}$ )	$\Phi$	$\tau$ (ns)	
b	In $\text{CH}_2\text{Cl}_2$ 	Exp.	501	-	-	513	-	0.63	3.97	
		Pred.	512	675	4.89	517	1186	0.59	3.7	
	In $\text{CH}_2\text{Cl}_2$ 	Exp.	506	-	-	517	-	0.072	-	
		Pred.	518	898	4.82	531	1156	0.097	2.0	
	In $\text{CH}_2\text{Cl}_2$ 	Exp.	506	815	4.78	522	1543	0.06	-	
		Pred.	521	960	4.89	537	1184	0.46	3.5	
	In $\text{CH}_2\text{Cl}_2$ 	Exp.	518	787	4.9	536	1141	0.635	6.9	
		Pred.	537	646	4.87	547	990	0.521	4.2	
	c	In water 	Exp.	342	-	-	407	-	-	2.3
			Pred.	362	4457	4.31	430	3154	0.103	2.0
		In water 	Exp.	332	-	-	398	-	-	7.69
			Pred.	342	4960	4.18	419	3586	0.17	2.5
In water 		Exp.	328	-	-	357	-	-	6.67	
		Pred.	311	4648	3.97	373	3787	0.244	15.2	

<sup>a</sup>(a) Various optical properties of chromophores with different molecular structures. (b) BODIPY derivatives (in red) with different moieties (in green) at the meso position in dichloromethane. (c) Three protonation states of 7-amino-2-naphthol in water and their experimentally measured and predicted values.

( $R^2$ ) ranges from 0.93 to 0.71 for the test sets (Table S2). It should be noted that  $\sigma_{\text{abs}}$ ,  $\sigma_{\text{emi}}$ ,  $\Phi$ , and  $\tau$  for a chromophore cannot be well predicted by currently available theoretical methods because they are highly dependent on many factors, including solute–solvent interactions, density of energy states,

and relaxation pathways, which cannot be fully considered in the theoretical methods. However, in our DL model, these properties are shown to be directly learned from the experimental database and reliably predicted, which is a great advantage of the DL method.

### Performances of Our DL Model

The DL model trained with the experimental database can be used to achieve DL optical spectroscopy such that seven optical properties for any given molecule are easily predicted. In this section, several interesting examples from the results of our DL model will be discussed.

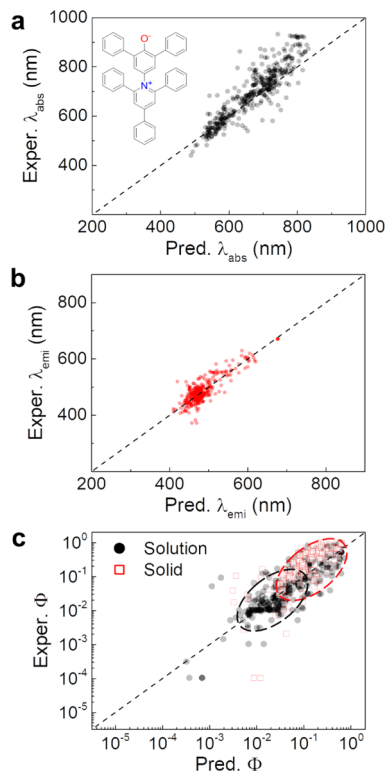
**Structural Diversity of Chromophores.** As shown in Table 1a, our DL model predicts the optical properties of widely used core structures of chromophores and fluorophores<sup>45–48</sup> such as Nile red,<sup>49</sup> Prodan,<sup>50</sup> methylene blue,<sup>51</sup> and L-tyrosine.<sup>52</sup> In addition, we compared the  $\lambda_{\text{abs}}$  and  $\lambda_{\text{emi}}$  values predicted by our DL model and TD-DFT calculations<sup>53</sup> for 131 combinations involving 53 chromophores (Figure S5 and Table S6). The RMSEs of  $\lambda_{\text{abs}}$  and  $\lambda_{\text{emi}}$  predicted by our model are 17.0 and 22.5 nm, respectively, which are 2–3.5 times smaller than those obtained by TD-DFT calculations (46.1 and 81.9 nm, respectively). The biggest advantage of DL method is that the information hidden in the experimental database is directly decoded by the DL model. As a result, the DL model can predict the properties (e.g.,  $\sigma_{\text{abs}}$ ,  $\sigma_{\text{emi}}$ ,  $\Phi$ , and  $\tau$ ) that are traditionally incomputable by theoretical methods. Compared to theory-based TD-DFT calculations, another advantage of the DL method is its low computational cost. The TD-DFT calculations required 236 min per molecule on average to compute the ground- and excited-state geometries,  $\lambda_{\text{abs}}$  and  $\lambda_{\text{emi}}$ . Our DL model required only 20 s to simultaneously predict the seven optical properties of 131 chromophore/solvent combinations on the same computer (see Methods for computational details).

For donor–acceptor (D–A) type molecules, the DL model is found to recognize the core moiety that has a more significant effect on the overall optical properties. Table 1b shows D–A type molecules with BODIPY and other moieties.<sup>54–58</sup> Our DL model confirms that the emission properties ( $\lambda_{\text{emi}}$ ) of the molecules are mainly determined by the BODIPY moiety. For example, for the anthracene (donor)–BODIPY (acceptor) structure, the emission peak position of BODIPY is longer than that of anthracene, and thus, BODIPY determines the emission peak position of the molecule according to Kasha's rule.<sup>59</sup>

In addition, we further examined whether the DL model could distinguish the optical properties of chromophores with similar structures. Ninety-seven chromophores in dichloromethane were selected from the experimental database (Table S7), and their optical properties were obtained by the DL model and compared with the experimental values (Figure S6). The DL model recognizes the core structure of a given molecule that determines the optical properties and the effect of structural modification of the core structures on the optical properties such as changes in the conjugation length and substituents of the core structures. The diverse changes in the core structures were successfully accounted for in the prediction of the optical properties.

Lastly, a change in protonation state is considered the smallest possible modification to a molecular structure which can dramatically alter its optical properties; this is because protonation/deprotonation changes the chromophore charge distribution and resultantly the chromophore–solvent interactions. Fifty-nine molecules with two or more protonation states were identified in the database (Figure S7). The effects of protonation/deprotonation in acid–base reaction on the optical properties are well predicted by the DL model. For example, the  $\lambda_{\text{abs}}$  and  $\lambda_{\text{emi}}$  of the three protonation states of 7-amino-2-naphthol in Table 1c are well distinguished.<sup>60</sup>

**Effects of Solvent Polarity.** The optical properties are significantly influenced by the solvent polarity known as solvatochromic shifts including bathochromic shift (red shift) and hypsochromic shift (blue shift). In the DL model, chromophore–solvent interactions are included, allowing reliable prediction of solvatochromic shifts. The solvatochromic shifts of Reichardt's dye<sup>61</sup> (Betaine 30, Figure 3a) in 334 solvent

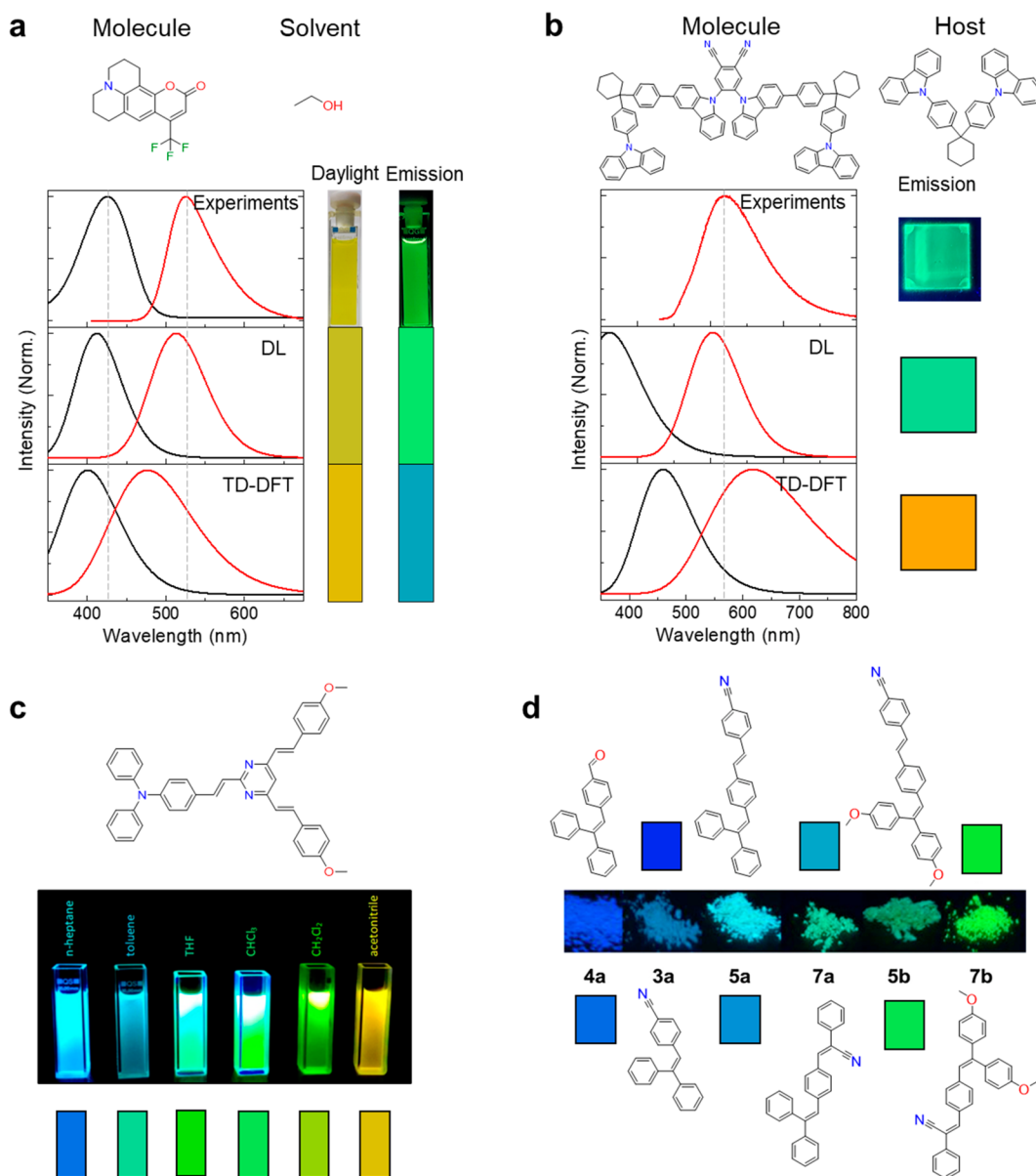


**Figure 3.** Solvent effects on optical properties predicted by our DL model. (a) Experimental vs predicted  $\lambda_{\text{abs}}$  values of Reichardt's dye (Betaine 30) in 334 solvents. The  $\lambda_{\text{abs}}$  of Reichardt's dye exhibits a hypsochromic (blue) shift from 1000 to 450 nm with increasing solvent polarity. (b) Experimental vs predicted  $\lambda_{\text{emi}}$  values of dopants in host matrices (films). (c) Experimental vs predicted  $\Phi$  values of molecules exhibiting aggregation induced emission in solutions or in solid states.

molecules are well predicted by the DL model. The DL model is found to quite accurately predict the  $\lambda_{\text{abs}}$  of Betaine 30 in solvents with similar dielectric constants ( $\epsilon_r$ ): 1,2-dichloroethane ( $\epsilon_r = 10.36$ ,  $\lambda_{\text{abs}}^{\text{exp}} = 692$  nm,  $\lambda_{\text{abs}}^{\text{pred}} = 692$  nm), 1,1-dichloroethane ( $\epsilon_r = 10.46$ ,  $\lambda_{\text{abs}}^{\text{exp}} = 726$  nm,  $\lambda_{\text{abs}}^{\text{pred}} = 701$  nm), and 2,2-dichloropropane ( $\epsilon_r = 11.63$ ,  $\lambda_{\text{abs}}^{\text{exp}} = 754$  nm,  $\lambda_{\text{abs}}^{\text{pred}} = 715$  nm). These results indicate that the chromophore–solvent interactions in the DL model were sufficiently well implemented.

Organic fluorophores can also be incorporated as dopants in solid-state host matrices. Such systems are widely used in OLEDs.<sup>62</sup> Our database includes 48 host molecules and 361 fluorophore(dopant)–host systems developed for OLEDs. The matrix effects on the emission properties of the dopants are well predicted (Figures 3b and S8). Estimation of the PLQY of dopants in host matrices is critical for designing dopant molecules, but the PLQYs of molecules in solutions or in films are challenging to estimate theoretically. The DL model could be effectively used to prescreen candidate dopant molecules by accurately estimating their PLQY.

In solid states, nonradiative processes associated with molecular rotations and vibrations are inactive and relatively



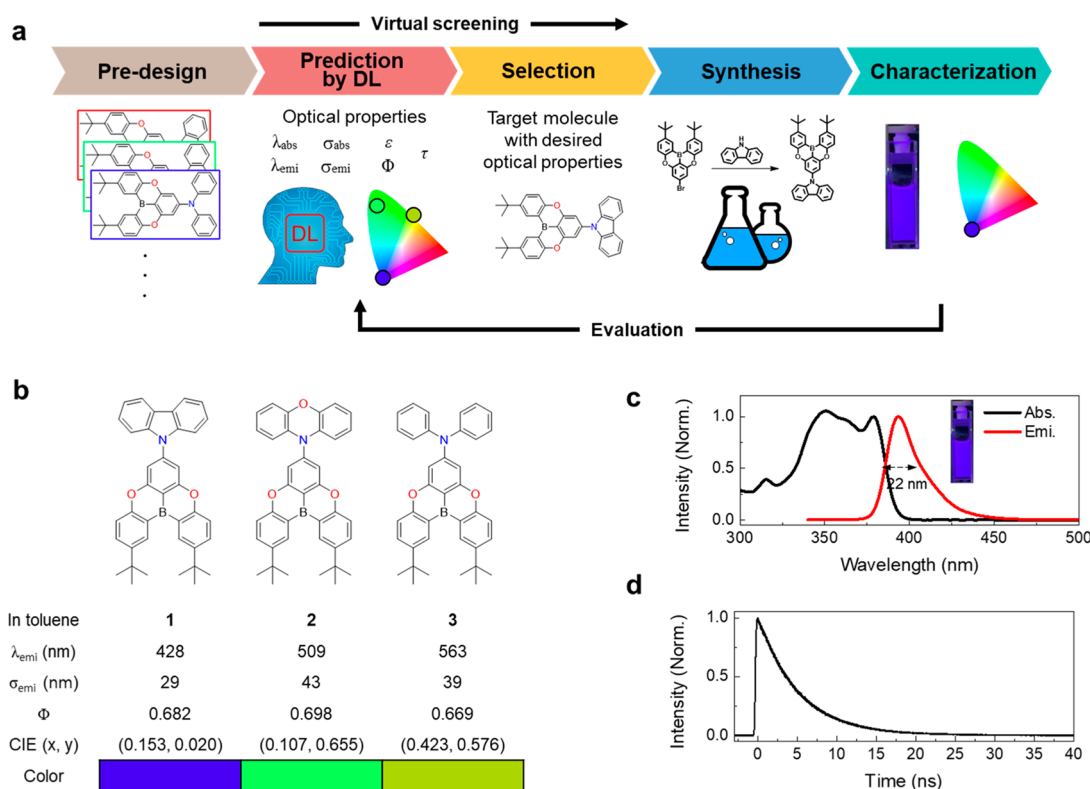
**Figure 4.** DL optical spectroscopy. Experimentally measured and DL predicted absorption (black) and emission (red) spectra. (a) Coumarin 153 in ethanol. (b) BPCP-2CPC (molecule) in C-2PC (host). The bandwidths in fwhm for the calculated spectra are set to  $5370\text{ cm}^{-1}$  (the default value in GaussView 6). (c) Photograph of *(E,E,E)*-2-(4-diphenylaminostyryl)-4,6-bis(4-methoxystyryl)pyrimidine in several solvents. The colors below the photograph are those predicted using our DL model [Reproduced with permission from ref 66. Copyright 2018 American Chemical Society]. (d) Photographs of solid state emission. The colors below the photograph are those predicted using our DL model [Reprinted with permission from ref 67. Copyright 2016 Published by Elsevier Ltd.].

high fluorescence quantum yields can be achieved. Aggregation-induced emission (AIE) has recently become an active research topic,<sup>63</sup> and we investigated whether our DL model could explain AIE. In our DL model, the effects of local environments are implemented using the adjacency and feature matrices of the solvents ( $A_{\text{sol}}$  and  $X_{\text{sol}}$ ), which can be treated as either the molecule itself for one-component solids (i.e., molecular solids) or as the host matrix for solid solutions. Therefore, the optical properties of chromophores in solids are able to be trained and predicted by the DL model. Note that amorphous and crystalline phase solid states are not distinguished in our database. Our database contains 730 examples of chromophores with AIE properties (Figure S9 and Table S3), and their  $\Phi$  are indeed predicted to be larger in aggregates (molecular solids) than in solutions as shown in Figure 3c. In short, the effects of

solvent and local environments on the optical properties of chromophores were successfully implemented to account for both solvatochromic shifts and the AIE effect.

### Deep Learning Optical Spectroscopy

For a molecule, absorption and emission properties, which are quantified using spectroscopic measurements, are characterized by their position ( $\lambda_{\text{abs}}/\lambda_{\text{emi}}$ ), bandwidth ( $\sigma_{\text{abs}}/\sigma_{\text{emi}}$ ), and amplitude ( $\epsilon/\Phi$ ). For any given organic compound, these optical properties are readily predicted by our DL model and can be used to simulate the absorption and emission spectra as described in Methods, making DL optical spectroscopy possible. Furthermore, the daylight and emission colors of a molecule in solution and films under the daylight and UV-lamp can be



**Figure 5.** Development of a target chromophore by virtual screening. (a) Overview for development of a new chromophore by DL optical spectroscopy. (b) Molecular structures of 3 molecules based on a DOBNA core with carbazole (1), phenoxazine (2), and diphenylamine moieties (3), and the optical properties predicted by DL optical spectroscopy. (c) Absorption and emission spectra of compound 1 in toluene.  $\sigma_{\text{abs}}$  cannot be experimentally measured because the  $S_1$  transition is overlapped with higher electronic transitions. (d) Time-resolved fluorescence signal of compound 1 in toluene.

estimated based on the absorption and emission spectra, respectively.

The absorption and emission spectra of coumarin 153 in ethanol were predicted using the DL model and TD-DFT calculations and are compared in Figure 4a.<sup>64</sup> The daylight and emission colors of coumarin 153 in ethanol are successfully reproduced by the DL model. In contrast, TD-DFT calculations estimated the daylight color well but not the emission color. For a molecule,  $\lambda_{\text{abs}}$ ,  $\lambda_{\text{emi}}$ ,  $\epsilon$ , and the radiative rate constant can be estimated by theoretical calculations, but  $\sigma_{\text{abs}}$ ,  $\sigma_{\text{emi}}$ , and  $\Phi$  are highly challenging to predict using currently available theoretical methods. The performance of the DL model was further examined using the thermally activated delayed fluorescence (TADF) dye, BPCP-2CPC, in the host molecule, C-2PC.<sup>65</sup> While the optical properties of BPCP-2CPC in toluene are included in our database, those of BPCP-2CPC in C-2PC are not. The emission spectrum and a photograph of BPCP-2CPC in C-2PC are presented in Figure 4b. The emission spectrum of BPCP-2CPC in C-2PC was reasonably well predicted by the DL model, and the emission color is well represented in the CIE 1931  $xy$  chromaticity space (Figure S10). Figures 4c, d and S11 show the photographs of chromophores in various environments along with the emission colors predicted by the DL model.<sup>66,67</sup> The solvent-polarity-dependent colors are well reproduced. Accordingly, our DL optical spectroscopy is successfully implemented to predict the optical properties of chromophores in solutions, films, and powders.

### Development of a Target Chromophore by Virtual Screening via DL Optical Spectroscopy

DL optical spectroscopy could be effectively used to select target chromophores with desired optical and photophysical properties from many pre-designed candidate molecules as shown in Figure 5a. Many potentially synthesizable molecular structures are pre-designed, their seven optical properties are predicted by DL optical spectroscopy, a target molecule with desired properties is selected and synthesized, and finally the optical and photophysical properties of the synthesized target molecule are characterized by many spectroscopic methods and whether the optical and photophysical properties match those desired is evaluated. In this way, new target molecules optimized for a given purpose can be efficiently developed using DL optical spectroscopy.

This section demonstrates how DL optical spectroscopy can be practically used in the development of a new blue fluorophore with a donor (D)–acceptor (A) type. Figure 5b shows three organic molecules with a DOBNA backbone which could be readily pre-designed. And then their optical properties in toluene were predicted by DL optical spectroscopy as summarized in Table S4. Depending on the donor unit such as carbazole (1), phenoxazine (2), and diphenylamine moieties (3) as shown in Figure 5b, the pre-designed molecules are predicted to exhibit quite different emission colors, very narrow emission bandwidths, and decent PLQYs. Note that DOBNA core structures were shown to exhibit such narrow bandwidths and high PLQYs.<sup>68–70</sup> Among three compounds in Figure 5b, Compound 1 is a fluorophore with a deep blue color ( $\lambda_{\text{emi}} = 428$  nm and CIE  $y < 0.1$ ), narrow emission bandwidth ( $\sigma_{\text{emi}} = 29$



nm), and decent PLQY ( $\Phi = \sim 0.7$ ). As a desired deep blue emitter, compound **1** was synthesized.

As shown in Figure 5c and d, the optical properties of compound **1** were experimentally measured in toluene, which is a typical solvent for optical measurements of organic compounds, and are summarized in Table S5. The optical properties of compound **1** in toluene are found to agree reasonably well with the values predicted by DL optical spectroscopy within the RMSEs although the experimentally determined values are somewhat slightly smaller than the predicted values. The  $\lambda_{\text{abs}}$  and  $\lambda_{\text{emi}}$  in the absorption and emission spectra are  $\sim 30$  nm blue-shifted compared to the predicted values. But the emission bandwidth ( $\sigma_{\text{emi}} = 22$  nm), PLQY ( $\Phi = 0.543$ ), and lifetime ( $\tau = 5.1$  ns) are reasonably well predicted. Furthermore, the CIE coordinates are found to be (0.157, 0.020), which is close to the predicted values (0.153, 0.020).

## CONCLUSIONS

We have successfully developed a DL method based on an experimental database to predict the optical properties of chromophores or fluorophores in various environments. Our DL model was trained using an experimental database of 30 094 chromophore/solvent combinations consisting of 11 392 unique organic chromophores in 369 solvents or solid states.<sup>11,12</sup> Seven optical properties of each organic molecule can be reliably estimated using the DL model and used to reconstruct the absorption and emission spectra and predict the real-world color of the molecule, achieving DL optical spectroscopy.

DL optical spectroscopy can be effectively used to prescreen whether newly designed chromophores and fluorophores exhibit desired optical properties before they are actually synthesized, which would reduce development costs significantly. We successfully demonstrated a strategy to developing a new deep blue emitter with desired optical properties using virtual screening by DL optical spectroscopy. By utilizing DL optical spectroscopy, the development of organic molecules for electronic and optoelectronic applications will be boosted.

Our DL method represents a framework that could be generally applicable to predict other chemical and physical properties of organic compounds. The experimental database of 30 094 chromophore/solvent combinations is currently in use but can be continuously extended to include a broader range of core structures. Therefore, the prediction accuracy of optical properties and the coverage of molecular structures by our DL optical spectroscopy will be continuously improved in the future. Recently, DL models that generate valid molecular structures based on a variational autoencoder (VAE) and a generative adversarial network (GAN) have been reported,<sup>71–76</sup> and by combining these generative DL models, our DL optical spectroscopy is expected to be used more efficiently in the development of new fluorophores and chromophores. Lastly, using our DL model we developed a web-based interface, “Deep4Chem”,<sup>77</sup> to predict the optical properties of organic compounds as well as the electronic absorption and emission peaks and the corresponding original colors which the materials would have, associated with absorption and emission peaks, in CIE color coordinates.

## METHODS

### Training Our DL Model

Our DL model was trained using an experimental database. In the training process, the loss function was defined by the total mean squared error which was obtained by summing the mean squared errors of seven optical and photophysical properties with the same weight. Note that  $\tau$  and  $\Phi$  were taken logarithms before normalization because their values reported in the literature were ranged on many orders of magnitude. In the case of  $\Phi = 0$  in the database,  $10^{-5}$  was added to  $\Phi$ . The absorption and emission bandwidths ( $\sigma_{\text{abs}}$  and  $\sigma_{\text{emi}}$ ) in nanometers in the database were converted to those in wavenumbers by using the following equation,  $\sigma_j(\text{cm}^{-1}) = 10^7/[\lambda_j - \sigma_j(\text{nm})/2] - 10^7/[\lambda_j + \sigma_j(\text{nm})/2]$  where  $\lambda_j$  is the absorption and emission peak positions in nanometers. In addition, all molecular properties were normalized to follow the standard normal distribution so that the loss of each property was on the same scale. A total of 30 094 chromophore/solvent combinations are randomly divided in a ratio of 0.8:0.1:0.1 for training, validation, and testing. For optimizing the hyper-parameters such as the number and dimension of GCN layers and MLPs, training and validation sets were used. No evidence of overfitting was found during the training over 10 000 epochs, as shown in Figure S3. After optimization, the DL model was trained by using training and validation sets. The training results are displayed in Figure S4 and Tables S1 and S2.

### Quantum Chemical Calculations

All quantum chemical calculations were performed using the density functional theory (DFT) and time-dependent DFT (TD-DFT) methods at the B3LYP level with a 6-31G(d) basis set as implemented in the Gaussian 16 package.<sup>53</sup> For all molecules in Table S5, the integral equation formalism for the polarizable continuum model (IEF-PCM) was used for solvation. The absorption and emission spectra of molecules were obtained by optimizing the electronic ground state and excited state, respectively, and subsequently calculating the TD-DFT transition energies. Using a computer with an Intel CPU (i7-6700) and 16 GB of RAM, TD-DFT calculations took an average of 236 min per molecule. On the other hand, our DL model took 20 s for 131 molecule–solvent pairs on the same computer.

### Reconstruction of Absorption and Emission Spectra

Absorption and emission spectra,  $S_j(\omega)$ , were calculated as a single peak using the normalized Gaussian function and predicted values by our DL model

$$S_j(\omega) = \exp\left[-4\ln(2)\left(\frac{10^7/\omega - 10^7/\lambda_j}{\sigma_j}\right)^2\right] \quad (1)$$

where  $\lambda_j$  is the absorption or emission peak position in nm and  $\sigma_j$  is the bandwidth. The vibronic features and asymmetric nature of the spectroscopic peaks were not reflected in eq 1. The absorption intensity is directly related to the extinction coefficient by Beer's law ( $A = \epsilon bc$ ); thus, the absorption spectrum can be approximated as  $\epsilon S_{\text{abs}}(\omega)$ . Similarly, the emission intensity is proportional to the absorption intensity and PLQY, and thus, the emission spectrum can be obtained as  $\epsilon\Phi S_{\text{emi}}(\omega)$ . The product of PLQY and extinction coefficient (i.e.,  $\Phi\epsilon$ ) is known as the brightness. Using the optical and photophysical properties ( $\lambda_{\text{abs}}$ ,  $\lambda_{\text{emi}}$ ,  $\sigma_{\text{abs}}$ ,  $\sigma_{\text{emi}}$ ,  $\epsilon$ , and  $\Phi$ ) predicted by the DL model, the absorption or emission spectra of a molecule were readily calculated. Furthermore, the CIE 1931 color space was computed using the calculated absorption and emission spectra. The CIE Standard Illuminant D65 was used for a daylight color.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/jacsau.1c00035>.

The details of our deep learning model, prediction by our deep learning model, and synthetic procedure of compound **1** (PDF)

## AUTHOR INFORMATION

### Corresponding Author

**Sungnam Park** – Department of Chemistry and Research Institute for Natural Science, Korea University, Seoul 02841, Korea; [orcid.org/0000-0001-6288-4620](https://orcid.org/0000-0001-6288-4620); Email: [spark8@korea.ac.kr](mailto:spark8@korea.ac.kr)

### Authors

**Joonyoung F. Joung** – Department of Chemistry and Research Institute for Natural Science, Korea University, Seoul 02841, Korea; [orcid.org/0000-0002-5976-2048](https://orcid.org/0000-0002-5976-2048)

**Minhi Han** – Department of Chemistry and Research Institute for Natural Science, Korea University, Seoul 02841, Korea; [orcid.org/0000-0002-9873-4904](https://orcid.org/0000-0002-9873-4904)

**Jinhyo Hwang** – Department of Chemistry and Research Institute for Natural Science, Korea University, Seoul 02841, Korea

**Minseok Jeong** – Department of Chemistry and Research Institute for Natural Science, Korea University, Seoul 02841, Korea

**Dong Hoon Choi** – Department of Chemistry and Research Institute for Natural Science, Korea University, Seoul 02841, Korea; [orcid.org/0000-0002-3165-0597](https://orcid.org/0000-0002-3165-0597)

Complete contact information is available at: <https://pubs.acs.org/10.1021/jacsau.1c00035>

### Author Contributions

<sup>†</sup>J.F.J. and M.H. contributed equally to this work.

### Author Contributions

S.P. and D.H.C. conceived the research. J.F.J., M.H., and M.J. conducted research under the supervision of S.P. J.F.J., M.H., and M.J. built the experimental database. J.H. synthesized and characterized compound **1**. J.F.J. and M.H. developed the deep learning algorithm and codes, and the web-based Deep4Chem. All authors analyzed the data. J.F.J., M.H., and S.P. wrote the manuscript.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This study was supported by grants from the National Research Foundation of Korea (NRF) funded by the Korean government (No. 2019R1A6A1A11044070) and Korea University-Future Research Grant (KU-FRG).

## REFERENCES

- (1) Burlingame, Q.; Huang, X.; Liu, X.; Jeong, C.; Coburn, C.; Forrest, S. R. Intrinsically stable organic solar cells under high-intensity illumination. *Nature* **2019**, *573* (7774), 394.
- (2) Uoyama, H.; Goushi, K.; Shizu, K.; Nomura, H.; Adachi, C. Highly efficient organic light-emitting diodes from delayed fluorescence. *Nature* **2012**, *492* (7428), 234.
- (3) Li, B.; Zhao, M.; Feng, L.; Dou, C.; Ding, S.; Zhou, G.; Lu, L.; Zhang, H.; Chen, F.; Li, X.; et al. Organic NIR-II molecule with long blood half-life for in vivo dynamic vascular imaging. *Nat. Commun.* **2020**, *11* (1), 3102.
- (4) Zhang, Y.; Yang, H.; Ma, H.; Bian, G.; Zang, Q.; Sun, J.; Zhang, C.; An, Z.; Wong, W. Y. Excitation Wavelength Dependent Fluorescence of

an ESIPT Triazole Derivative for Amine Sensing and Anti-Counterfeiting Applications. *Angew. Chem.* **2019**, *131* (26), 8865.

(5) Woodward, R. B. Structure and the Absorption Spectra of  $\alpha,\beta$ -Unsaturated Ketones. *J. Am. Chem. Soc.* **1941**, *63* (4), 1123.

(6) Roothaan, C. C. J. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **1951**, *23* (2), 69.

(7) McWeeny, R.; Diercksen, G. Self-Consistent Perturbation Theory. II. Extension to Open Shells. *J. Chem. Phys.* **1968**, *49* (11), 4852.

(8) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136* (3B), B864.

(9) Bauernschmitt, R.; Ahlrichs, R. Treatment of electronic excitations within the adiabatic approximation of time dependent density functional theory. *Chem. Phys. Lett.* **1996**, *256* (4–5), 454.

(10) Adamo, C.; Jacquemin, D. The calculations of excited-state properties with Time-Dependent Density Functional Theory. *Chem. Soc. Rev.* **2013**, *42* (3), 845.

(11) Joung, J. F.; Han, M.; Jeong, M.; Park, S. *Figshare* **2020**, DOI: [10.6084/m9.figshare.12045567.v2](https://doi.org/10.6084/m9.figshare.12045567.v2).

(12) Joung, J. F.; Han, M.; Jeong, M.; Park, S. Experimental database of optical properties of organic compounds. *Sci. Data* **2020**, *7* (1), 295.

(13) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559* (7715), 547.

(14) Mater, A. C.; Coote, M. L. Deep Learning in Chemistry. *J. Chem. Inf. Model.* **2019**, *59* (6), 2545.

(15) Montavon, G.; Rupp, M.; Gobbi, V.; Vazquez-Mayagoitia, A.; Hansen, K.; Tkatchenko, A.; Müller, K.-R.; Anatole von Lilienfeld, O. Machine learning of molecular electronic properties in chemical compound space. *New J. Phys.* **2013**, *15* (9), 095003.

(16) Coley, C. W.; Barzilay, R.; Green, W. H.; Jaakkola, T. S.; Jensen, K. F. Convolutional Embedding of Attributed Molecular Graphs for Physical Property Prediction. *J. Chem. Inf. Model.* **2017**, *57* (8), 1757.

(17) Musil, F.; De, S.; Yang, J.; Campbell, J. E.; Day, G. M.; Ceriotti, M. Machine learning for the structure-energy-property landscapes of molecular crystals. *Chem. Sci.* **2018**, *9* (5), 1289.

(18) Jinich, A.; Sanchez-Lengeling, B.; Ren, H.; Harman, R.; Aspuru-Guzik, A. A Mixed Quantum Chemistry/Machine Learning Approach for the Fast and Accurate Prediction of Biochemical Redox Potentials and Its Large-Scale Application to 315000 Redox Reactions. *ACS Cent. Sci.* **2019**, *5* (7), 1199.

(19) Yamada, H.; Liu, C.; Wu, S.; Koyama, Y.; Ju, S.; Shiomi, J.; Morikawa, J.; Yoshida, R. Predicting Materials Properties with Little Data Using Shotgun Transfer Learning. *ACS Cent. Sci.* **2019**, *5* (10), 1717.

(20) Afzal, M. A. F.; Sonpal, A.; Haghightalari, M.; Schultz, A. J.; Hachmann, J. A deep neural network model for packing density predictions and its application in the study of 1.5 million organic molecules. *Chem. Sci.* **2019**, *10* (36), 8374.

(21) Dong, Y.; Wu, C.; Zhang, C.; Liu, Y.; Cheng, J.; Lin, J. Bandgap prediction by deep learning in configurationally hybridized graphene and boron nitride. *npj Comput. Mater.* **2019**, *5* (1), 26.

(22) So, S.; Mun, J.; Rho, J. Simultaneous Inverse Design of Materials and Structures via Deep Learning: Demonstration of Dipole Resonance Engineering Using Core-Shell Nanoparticles. *ACS Appl. Mater. Interfaces* **2019**, *11* (27), 24264.

(23) Ye, S.; Hu, W.; Li, X.; Zhang, J.; Zhong, K.; Zhang, G.; Luo, Y.; Mukamel, S.; Jiang, J. A neural network protocol for electronic excitations of N-methylacetamide. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (24), 11612.

(24) Ghosh, K.; Stuke, A.; Todorovic, M.; Jorgensen, P. B.; Schmidt, M. N.; Vehtari, A.; Rinke, P. Deep Learning Spectroscopy: Neural Networks for Molecular Excitation Spectra. *Adv. Sci.* **2019**, *6* (9), 1801367.

(25) Back, S.; Yoon, J.; Tian, N.; Zhong, W.; Tran, K.; Ulissi, Z. W. Convolutional Neural Network of Atomic Surface Structures To Predict Binding Energies for High-Throughput Screening of Catalysts. *J. Phys. Chem. Lett.* **2019**, *10* (15), 4401.

- (26) Kang, B.; Seok, C.; Lee, J. Prediction of Molecular Electronic Transitions Using Random Forests. *J. Chem. Inf. Model.* **2020**, *60* (12), 5984.
- (27) Na, G. S.; Jang, S.; Lee, Y. L.; Chang, H. Tuplewise Material Representation Based Machine Learning for Accurate Band Gap Prediction. *J. Phys. Chem. A* **2020**, *124* (50), 10616.
- (28) Ye, S.; Zhong, K.; Zhang, J.; Hu, W.; Hirst, J. D.; Zhang, G.; Mukamel, S.; Jiang, J. A Machine Learning Protocol for Predicting Protein Infrared Spectra. *J. Am. Chem. Soc.* **2020**, *142* (45), 19071.
- (29) Li, B.; Liu, R.; Bai, H.; Ju, C.-W. Can Machine Learning Be More Accurate Than TD-DFT? Prediction of Emission Wavelengths and Quantum Yields of Organic Fluorescent Materials. *ChemRxiv* **2020**, DOI: 10.26434/chemrxiv.12111060.v1.
- (30) Gao, H.; Struble, T. J.; Coley, C. W.; Wang, Y.; Green, W. H.; Jensen, K. F. Using Machine Learning To Predict Suitable Conditions for Organic Reactions. *ACS Cent. Sci.* **2018**, *4* (11), 1465.
- (31) Segler, M. H. S.; Preuss, M.; Waller, M. P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **2018**, *555* (7698), 604.
- (32) Voznyy, O.; Levina, L.; Fan, J. Z.; Askerka, M.; Jain, A.; Choi, M. J.; Ouellette, O.; Todorovic, P.; Sagar, L. K.; Sargent, E. H. Machine Learning Accelerates Discovery of Optimal Colloidal Quantum Dot Synthesis. *ACS Nano* **2019**, *13* (10), 11122.
- (33) Torng, W.; Altman, R. B. Graph Convolutional Neural Networks for Predicting Drug-Target Interactions. *J. Chem. Inf. Model.* **2019**, *59* (10), 4131.
- (34) Lim, J.; Hwang, S.-Y.; Moon, S.; Kim, S.; Kim, W. Y. Scaffold-based molecular design with a graph generative model. *Chem. Sci.* **2020**, *11* (4), 1153.
- (35) Jha, D.; Ward, L.; Paul, A.; Liao, W. K.; Choudhary, A.; Wolverton, C.; Agrawal, A. ElemNet: Deep Learning the Chemistry of Materials From Only Elemental Composition. *Sci. Rep.* **2018**, *8* (1), 17593.
- (36) Sahu, H.; Yang, F.; Ye, X.; Ma, J.; Fang, W.; Ma, H. Designing promising molecules for organic solar cells via machine learning assisted virtual screening. *J. Mater. Chem. A* **2019**, *7* (29), 17480.
- (37) Kim, B.; Lee, S.; Kim, J. Inverse design of porous materials using artificial neural networks. *Sci. Adv.* **2020**, *6* (1), eaax9324.
- (38) Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; von Lilienfeld, O. A.; Muller, K. R.; Tkatchenko, A. Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *J. Phys. Chem. Lett.* **2015**, *6* (12), 2326.
- (39) Hermann, J.; Schätzle, Z.; Noé, F. Deep-neural-network solution of the electronic Schrödinger equation. *Nat. Chem.* **2020**, *12* (10), 891.
- (40) Nakata, M.; Shimazaki, T. PubChemQC Project: A Large-Scale First-Principles Electronic Structure Database for Data-Driven Chemistry. *J. Chem. Inf. Model.* **2017**, *57* (6), 1300.
- (41) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural Message Passing for Quantum Chemistry. *arXiv.org* **2017**, 1704.01212v2.
- (42) Ryu, S.; Lim, J.; Hong, S. H.; Kim, W. Y. Deeply learning molecular structure-property relationships using attention- and gate-augmented graph convolutional network. *arXiv.org* **2018**, 1805.10988.
- (43) Wang, X.; Li, Z.; Jiang, M.; Wang, S.; Zhang, S.; Wei, Z. Molecule Property Prediction Based on Spatial Graph Embedding. *J. Chem. Inf. Model.* **2019**, *59* (9), 3817.
- (44) Li, X.; Yan, X.; Gu, Q.; Zhou, H.; Wu, D.; Xu, J. DeepChemStable: Chemical Stability Prediction with an Attention-Based Graph Convolution Network. *J. Chem. Inf. Model.* **2019**, *59* (3), 1044.
- (45) Joung, J. F.; Kim, S.; Park, S. Cationic Effect on the Equilibria and Kinetics of the Excited-State Proton Transfer Reaction of a Photoacid in Aqueous Solutions. *J. Phys. Chem. B* **2018**, *122* (19), 5087.
- (46) Niko, Y.; Hiroshige, Y.; Kawauchi, S.; Konishi, G.-i. Fundamental photoluminescence properties of pyrene carbonyl compounds through absolute fluorescence quantum yield measurement and density functional theory. *Tetrahedron* **2012**, *68* (31), 6177.
- (47) Ciubini, B.; Visentin, S.; Serpe, L.; Canaparo, R.; Fin, A.; Barbero, N. Design and synthesis of symmetrical pentamethine cyanine dyes as NIR photosensitizers for PDT. *Dyes Pigm.* **2019**, *160*, 806.
- (48) Khopkar, S.; Jachak, M.; Shankarling, G. Novel A(2)-D-A(1)-D-A(2) type NIR absorbing symmetrical squaraines based on 2, 3, 3, 8-tetramethyl-3H-pyrrolo [3, 2-h] quinoline: Synthesis, photophysical, electrochemical, thermal properties and photostability study. *Spectrochim. Acta, Part A* **2019**, *211*, 114.
- (49) Cser, A.; Nagy, K.; Biczók, L. Fluorescence lifetime of Nile Red as a probe for the hydrogen bonding strength with its microenvironment. *Chem. Phys. Lett.* **2002**, *360* (5–6), 473.
- (50) Niko, Y.; Kawauchi, S.; Konishi, G. Solvatochromic pyrene analogues of Prodan exhibiting extremely high fluorescence quantum yields in apolar and polar solvents. *Chem. - Eur. J.* **2013**, *19* (30), 9760.
- (51) Santin, L. R. R.; dos Santos, S. C.; Novo, D. L. R.; Bianchini, D.; Gerola, A. P.; Braga, G.; Caetano, W.; Moreira, L. M.; Bastos, E. L.; Romani, A. P.; et al. Study between solvatochromism and steady-state and time-resolved fluorescence measurements of the Methylene blue in binary mixtures. *Dyes Pigm.* **2015**, *119*, 12.
- (52) Wunsch, U. J.; Murphy, K. R.; Stedmon, C. A. Fluorescence Quantum Yields of Natural Organic Matter and Organic Compounds: Implications for the Fluorescence-based Interpretation of Organic Matter Composition. *Front. Mar. Sci.* **2015**, DOI: 10.3389/fmars.2015.00098.
- (53) 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.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16*; Gaussian, Inc.: Wallingford, CT, 2016.
- (54) Sun, H.; Dong, X.; Liu, S.; Zhao, Q.; Mou, X.; Yang, H. Y.; Huang, W. Excellent BODIPY Dye Containing Dimesitylboryl Groups as PeT-Based Fluorescent Probes for Fluoride. *J. Phys. Chem. C* **2011**, *115* (40), 19947.
- (55) Li, Z.; Lv, X.; Chen, Y.; Fu, W.-F. Synthesis, structures and photophysical properties of boron-fluorine derivatives based on pyridine/1,8-naphthyridine. *Dyes Pigm.* **2014**, *105*, 157.
- (56) Zhang, S.; Liu, X.; Yuan, W.; Zheng, W.; Li, H.; Li, C.; Sun, Y.; Wang, Y.; Yang, Y.; Li, Y.; et al. New aryl substituted pyridylimidazo-[1,2-a]pyridine-BODIPY conjugates: Emission color tuning from green to NIR. *Dyes Pigm.* **2018**, *159*, 406.
- (57) Filatov, M. A.; Karuthedath, S.; Polestshuk, P. M.; Callaghan, S.; Flanagan, K. J.; Telitchko, M.; Wiesner, T.; Laquai, F.; Senge, M. O. Control of triplet state generation in heavy atom-free BODIPY-anthracene dyads by media polarity and structural factors. *Phys. Chem. Chem. Phys.* **2018**, *20* (12), 8016.
- (58) Zhang, X. F.; Zhang, G. Q.; Zhu, J. Methylated Unsymmetric BODIPY Compounds: Synthesis, High Fluorescence Quantum Yield and Long Fluorescence Time. *J. Fluoresc.* **2019**, *29* (2), 407.
- (59) Kasha, M. Characterization of electronic transitions in complex molecules. *Discuss. Faraday Soc.* **1950**, *9*, 14.
- (60) Cotter, L. F.; Brown, P. J.; Nelson, R. C.; Takematsu, K. Divergent Hammett Plots of the Ground- and Excited-State Proton Transfer Reactions of 7-Substituted-2-Naphthol Compounds. *J. Phys. Chem. B* **2019**, *123* (19), 4301.
- (61) Reichardt, C. Solvatochromic Dyes as Solvent Polarity Indicators. *Chem. Rev.* **1994**, *94* (8), 2319.
- (62) Nakanotani, H.; Higuchi, T.; Furukawa, T.; Masui, K.; Morimoto, K.; Numata, M.; Tanaka, H.; Sagara, Y.; Yasuda, T.;

Adachi, C. High-efficiency organic light-emitting diodes with fluorescent emitters. *Nat. Commun.* **2014**, *5*, 4016.

(63) Hong, Y.; Lam, J. W.; Tang, B. Z. Aggregation-induced emission. *Chem. Soc. Rev.* **2011**, *40* (11), 5361.

(64) Jones, G.; Jackson, W. R.; Choi, C. Y.; Bergmark, W. R. Solvent effects on emission yield and lifetime for coumarin laser dyes. Requirements for a rotatory decay mechanism. *J. Phys. Chem.* **1985**, *89* (2), 294.

(65) Kim, H. J.; Kim, S. K.; Godumala, M.; Yoon, J.; Kim, C. Y.; Jeong, J. E.; Woo, H. Y.; Kwon, J. H.; Cho, M. J.; Choi, D. H. Novel molecular triad exhibiting aggregation-induced emission and thermally activated fluorescence for efficient non-doped organic light-emitting diodes. *Chem. Commun.* **2019**, *55* (64), 9475.

(66) Feckova, M.; le Poul, P.; Guen, F. R.; Roisnel, T.; Pytela, O.; Klikar, M.; Bures, F.; Achelle, S. 2,4-Distyryl- and 2,4,6-Tristyrylpyrimidines: Synthesis and Photophysical Properties. *J. Org. Chem.* **2018**, *83* (19), 11712.

(67) Wen, W.; Shi, Z.-F.; Cao, X.-P.; Xu, N.-S. Triphenylethylene-based fluorophores: Facile preparation and full-color emission in both solution and solid states. *Dyes Pigm.* **2016**, *132*, 282.

(68) Hirai, H.; Nakajima, K.; Nakatsuka, S.; Shiren, K.; Ni, J.; Nomura, S.; Ikuta, T.; Hatakeyama, T. One-Step Borylation of 1,3-Diaryloxybenzenes Towards Efficient Materials for Organic Light-Emitting Diodes. *Angew. Chem., Int. Ed.* **2015**, *54* (46), 13581.

(69) Kondo, Y.; Yoshiura, K.; Kitera, S.; Nishi, H.; Oda, S.; Gotoh, H.; Sasada, Y.; Yanai, M.; Hatakeyama, T. Narrowband deep-blue organic light-emitting diode featuring an organoboron-based emitter. *Nat. Photonics* **2019**, *13* (10), 678.

(70) Madayanad Suresh, S.; Hall, D.; Beljonne, D.; Olivier, Y.; Zysman-Colman, E. Multiresonant Thermally Activated Delayed Fluorescence Emitters Based on Heteroatom-Doped Nanographenes: Recent Advances and Prospects for Organic Light-Emitting Diodes. *Adv. Funct. Mater.* **2020**, *30* (33), 1908677.

(71) Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **2018**, *361* (6400), 360.

(72) Popova, M.; Isayev, O.; Tropsha, A. Deep reinforcement learning for de novo drug design. *Sci. Adv.* **2018**, *4* (7), eaap7885.

(73) Lim, J.; Ryu, S.; Kim, J. W.; Kim, W. Y. Molecular generative model based on conditional variational autoencoder for de novo molecular design. *J. Cheminf.* **2018**, *10* (1), 31.

(74) Elton, D. C.; Boukouvalas, Z.; Fuge, M. D.; Chung, P. W. Deep learning for molecular design—a review of the state of the art. *Mol. Syst. Des. Eng.* **2019**, *4* (4), 828.

(75) Hong, S. H.; Ryu, S.; Lim, J.; Kim, W. Y. Molecular Generative Model Based on an Adversarially Regularized Autoencoder. *J. Chem. Inf. Model.* **2020**, *60* (1), 29.

(76) Krenn, M.; Häse, F.; Nigam, A.; Friederich, P.; Aspuru-Guzik, A. Self-referencing embedded strings (SELFIES): A 100% robust molecular string representation. *Mach. Learn.: Sci. Technol.* **2020**, *1* (4), 045024.

(77) Joung, J. F.; Han, M.; Park, S. Deep4Chem, 2020; <http://deep4chem.korea.ac.kr/>.