ADVANCED OPTICAL MATERIALS
www.advopticalmat.de

Automatic Screen-out of Ir(III) Complex Emitters by Combined Machine Learning and Computational Analysis

Zheng Cheng, Jiapeng Liu, Tong Jiang, Mohan Chen, Fuzhi Dai, Zhifeng Gao, Guolin Ke, Zifeng Zhao, and Qi Ou**

The organic light-emitting diode (OLED) has gained widespread commercial use, yet there is a continuous need to identify innovative emitters that offer higher efficiency and a broader color gamut. To effectively screen out promising OLED molecules that are yet to be synthesized, representation learning aided high throughput virtual screening (HTVS) over millions of Ir(III) complexes, which are prototypical types of phosphorescent OLED material constructed via a random combination of 278 reported ligands. This study successfully screens out a decent amount of promising candidates for both display and lighting purposes, which are worth further experimental investigation. The high efficiency and accuracy of this model are largely attributed to the pioneering attempt of using representation learning to organic luminescent molecules, which is initiated by a pre-training procedure with over 1.6 million 3D molecular structures and frontier orbital energies predicted via semi-empirical methods, followed by a fine-tuning scheme via the quantum mechanical computed properties over around 1500 candidates. Such workflow enables an effective model construction process that is otherwise hindered by the scarcity of labeled data and can be straightforwardly extended to the discovery of other novel materials.

1. Introduction

First demonstrated by C. W. Tang and S. A. Vanslyke in 1987,[1] organic light-emitting diodes (OLEDs) have gradually become a mainstream display technology in consumer electronics due to their superior color properties and the capability of being made flexible.[2] Despite the fact that OLEDs have been commercialized for years, challenges still remain in terms of higher efficiency and wider color gamut, which require the discovery of novel molecules with desired luminescence properties.[3–7] The conventional way to design new materials relies on expertise intuition and sufficient amount of experimental validations, of which the trial-and-error cost is usually burdensome.

The rapid development of massive computational resources with advanced simulation and theoretical algorithms has made high-throughput virtual screening (HTVS) a ground-breaking tool in the design of new materials.[8–11] By comprehensively exploring the chemical space with tailored properties, HTVS succeeds in predicting the most promising candidates for experimental validation, and the trial-and-error cost can be remarkably reduced. Recent applications of this approach include the search for both organic and inorganic materials in the field of batteries,[12–17] 2D materials,[18–20] alloys,[21,22] semiconductors,[23,24] catalyst,[25,26] light-emitting devices,[27,28] and photovoltaics,[29,30] etc.[31–34] The central part of HTVS is the screening criteria, of which the generation relies on sufficient existing experimental data and/or accurate yet efficient quantum mechanical (QM) calculations. Unfortunately, experimental data under consistent conditions are rather limited, and highly

Z. Cheng
School of Mathematical Sciences
Peking University
Beijing 100871, P. R. China
Z. Cheng, J. Liu, T. Jiang, F. Dai, Z. Zhao, Q. Ou
AI for Science Institute
Beijing 100084, P. R. China
E-mail: zhaozf@bjaisi.com; ouq@bjaisi.com
J. Liu
School of Advanced Energy
Sun Yat-Sen University
Shenzhen 518107, P. R. China
T. Jiang
MOE Key Laboratory of Organic OptoElectronics and Molecular Engineering
Department of Chemistry
Tsinghua University
Beijing 100084, P. R. China
M. Chen
HEDPS, CAPT, College of Engineering and School of Physics
Peking University
Beijing 100871, P. R. China
Z. Gao, G. Ke
DP Technology
Beijing 100080, P. R. China

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adom.202301093

DOI: 10.1002/adom.202301093

accurate QM calculations are usually computationally demanding and even prohibitive to large realistic systems.[35–37] Both of these experimental and theoretical obstacles pose practical challenges to the application of HTVS.

By constructing quantitative structure-activity relationships (QSAR), machine learning (ML) algorithm can be applied as a powerful tool to efficiently predict experimental or QM properties and remarkably alleviate the data generation cost. Early property-predicting works of ML algorithms employ manually constructed features for the description of the structures, which requires case-by-case design for different systems.[38] Graph convolution neural networks provide a general scheme by representing the structure via neural network layers with atomic information and positions of systems as initial input[39], which gains pronounced improvement in terms of generalizability and transferability. The essential issue of ML-accelerated HTVS is indeed to achieve a desired accuracy and transferability with as little data as possible. Recently, attempts have been made to improve the data-efficiency by employing a more information-rich and reliable representation to better describe the system structures, among which E(3)-equivariant graph neural network (E3NN) is a good example.[40] Other studies employ the transfer learning scheme, in which the ML models is pre-trained with tremendous inexpensive labeled data and then fine-tuned with limited expensive but accurate labels.[41–43]

In 2022, we proposed Uni-Mol, a molecular representation learning scheme (MRL)[44–48], to further improve the data-efficiency by enabling the utilization of *unlabeled* data as prior knowledge. Within the framework of Uni-Mol, the property-predicting model is constructed with a pre-training process of tremendous unlabeled 3D structure data and cost-effective labeled data, followed by the application of fine-tuning schema with limited expensive labeled data on multiple downstream tasks.[49] Besides utilizing both labeled and unlabeled data for pre-training, Uni-Mol also employs a pair-level representation to obtain information-rich description of structures, making it a competitive tool in predicting structural-related properties of various materials.[49] In fact, by taking the advantage of enormous 3D conformations of the target systems and the state-of-the-art (SOTA) neural network, Uni-Mol rivals the performances of some advanced methods such as D-MPNN and AttentiveFP in predicting energetic properties of open-source organic and medical molecules.[49]

In this work, we expand the application of Uni-Mol to the HTVS of organic luminescent molecules, of which the light-emitting properties are closely related to their 3D conformations. We demonstrate that with the accurately benchmarked QM calculations on a limited amount of systems combined with the Uni-Mol training process over millions of automatically constructed and semi-empirically optimized structures, photophysical properties of Ir(III) complexes, which are prototypical phosphorescent molecules, can be efficiently predicted, enabling the screen-out of presumably outstanding candidates. Most importantly, such computational protocol can be effortlessly transferred to other organic materials such as thermally activated delayed fluorescence (TADF) molecules and organic photovoltaic, facilitating an efficient means of material design in a variety of fields.

2. Experimental Section

The overall workflow of HTVS process is schematized in **Figure 1**. With 278 bidentate ligands collected from previously published experimental studies on Ir(III) complex,[50] millions of candidate molecules were constructed using stk[51] and further optimized via density functional tight binding approach, i.e., GFN2-xTB.[52] A structure rationality filter was then introduced to filter out unphysical structures based on atomic distances and/or ligand angles. As such, around 1.6 million candidates were constructed, of which the coordinates and symbols were served as as the initial guesses for QM calculations and self-learning label during the pre-training process of Uni-Mol, while the energies of highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) at GFN2-xTB level were served as supervise-learning labels. Details of the framework, the pre-training and fine-tuning processes, and the optimization of the hyperparameters (Table S1, Supporting Information) of Uni-Mol were illustrated in the Supporting Information.

Due to the high computational cost of QM calculations, they were only carried out on 1468 randomly selected molecules with less than 61 heavy atoms from the candidate pool. The number of heavy atoms of the whole candidate pool ranges from 22 to 157, and the overall distribution could be seen in Figure S3 (Supporting Information). Four QM predicted properties, HOMO, LUMO, the adiabatic excitation energy of T_1 (E_{ad}), and the photoluminescence quantum yield (PLQY), were utilized to fine-tune the pre-trained model. All electronic structure calculations were performed with quantum chemistry packages Gaussian16[53] and ORCA,[54,55] while the photophysical properties including emission spectra and various rate constants were calculated via the thermal vibration correlation function (TVCF) method in the molecular material property prediction package MOMAP.[56–58]

It should be noted all properties were calculated for isolated gaseous phase molecules. In fact, such single-molecule model had been widely employed to analyze both the molecular orbitals and the excitation energies of solid-state organic light-emitting systems.[59–61] The rationality of such single-molecule model lied in the fact that for most OLED devices, luminescent molecules were doped on thin films, and such amorphous pattern usually limits the intermolecular interactions, making the single-molecule model a sound approximation to the practical scenario. The accuracy of the QM calculation process was validated by comparing the theoretical predicted adiabatic excitation energy and PLQY with the corresponding experimental values on ten selected molecules (as listed in **Table 1**). The molecular structures of these ten molecules were provided in Figure S4 (Supporting Information).

After the pre-training and fine-tuning process, the established Uni-Mol model was applied to single out potentially outstanding candidates, followed by further QM validation on these molecules. Target molecules for displaying and lighting purposes were then screened out with corresponding criteria (as detailed in the next section). Additional details of the HTVS workflow (Figure S2, Supporting Information) and the applied electronic structure theory were provided in the Supporting Information.

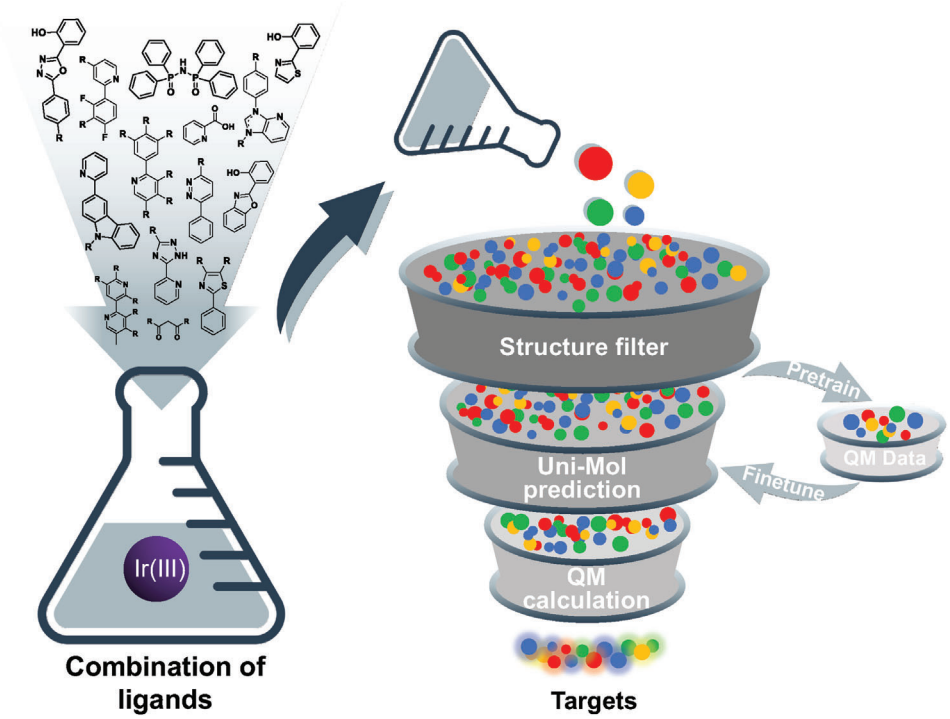

Figure 1. Schematic HTVS workflow for Ir(III) complex emitters.

Table 1. PLQY and adiabatic excitation energy (E_{ad}) of ten OLED molecules predicted by QM calculation and the corresponding experimental values.

Molecule ID	PLQY		E_{ad} (eV)	
	QM	Exp.	QM	Exp.[a)]
$(ppy)_2Ir(oz)$[62]	0.30	0.55	2.60	2.58
$Ir(dpt)_3$[3]	0.99	0.64	3.14	3.13
Complex 2[63]	0.39	0.35	2.62	2.44
$(mdp)_2Ir(acac)$[64]	0.76	0.85	2.40	2.37
Complex 1[65]	0.78	0.78	2.97	3.12
IrS-5F[66]	0.67	0.95	2.47	2.35
Ir1[67]	0.99	0.93	2.46	2.75
2FBNO[68]	0.83	0.71	2.37	2.45
Complex 2[69]	0.90	0.73	2.63	2.94
Ir5b[70]	0.99	0.69	2.91	2.96

[a)] The experiment values are estimated from the average of the absorption and emission energies.

3. Results and Discussion

3.1. Performance of Uni-Mol on Open-Source OLED Dataset

The capability of Uni-Mol in predicting electronic structure properties for organic optical molecules is first validated on an open-source solvated organic fluorescent dyes dataset,[71] of which the 3D information required by Uni-Mol training is generated from 2D simplified molecular-input line-entry system (SMILES) via Rdkit. Note that the solvent effect is not taken into consideration in our training process and hence molecules solvated with different solvents are removed. The scaffold splitting is applied to divide the dataset into training, validation, and test sets in the ratio of 8:1:1. As shown in **Figure 2**, our Uni-Mol model is able to provide accurate ML predictions for absorption and emission wavelengths. The correlation coefficient (R) and the MAE on test set are 0.991 and 7.3 nm for the absorption wavelength and 0.960 and 16.4 nm for the emission wavelength, which outperforms the previously reported results.[72] Therefore, Uni-Mol model can be rationally applied to our constructed candidate pool to initiate an accurate and cost-effective approach that connects 3D information of molecules and their optical properties.

3.2. Uni-Mol Performance on Ir(III) Complex Emitters

Next, we examine the performance of Uni-Mol on the constructed Ir(III) complexes pool. The pre-training process of Uni-Mol is performed over the whole candidate pool, with the energy of the frontier orbitals at GFN2-xTB level as labels. By dividing the dataset into training, validation, and test sets with the ratio of 8:1:1, the R value for HOMO and LUMO on the test set is 0.995 and 0.934, respectively, which evinces the consistency between Uni-Mol prediction and GFN2-xTB calculations. Such pre-training process is followed by the fine-tuning with four QM predicted properties, i.e., HOMO, LUMO, adiabatic excitation energy of T_1, and PLQY. The ratio of training, validation and test set is also set to 8:1:1.

The performance of the established Uni-Mol model on predicting these four optical properties is shown in **Figure 3**. It can be seen that highly accurate predictions of HOMO (MAE = 0.067 eV, R = 0.95), LUMO (MAE = 0.114 eV, R = 0.87), and E_{ad} (MAE

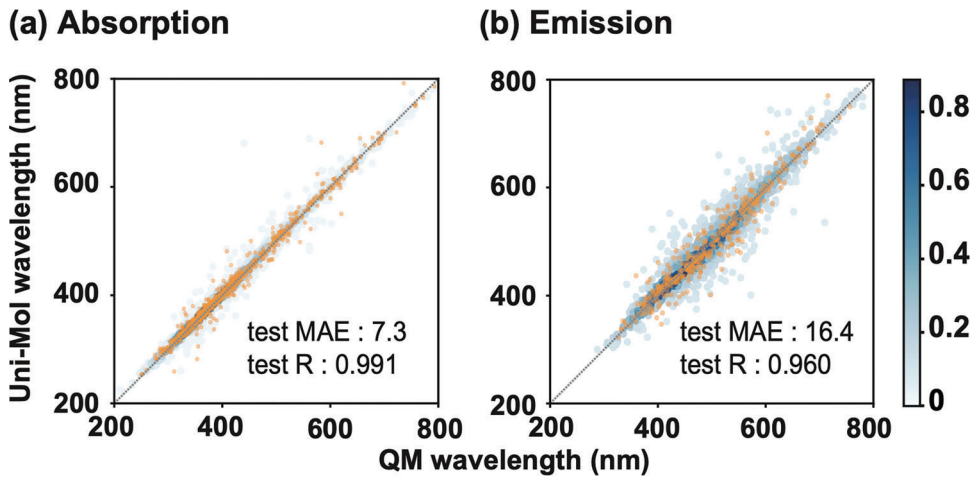

Figure 2. Correlation between the Uni-Mol predicted values [a) absorption and b) emission wavelength] and the experimental values on an open-source database.[71] The blue and yellow dots represent the training and test set, respectively. The color bar represents the density of data points in the absorption and emission training set.

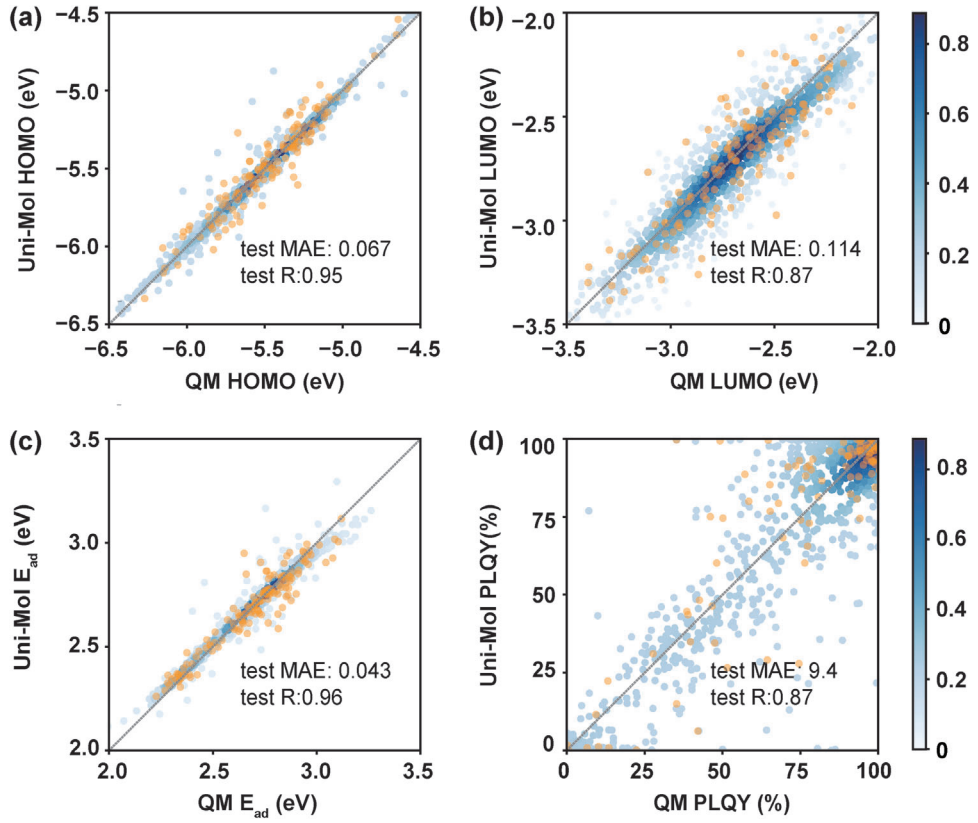

Figure 3. Correlation between the Uni-Mol predicted values [a) HOMO, b) LUMO, c) adiabatic excitation energy, and d) PLQY] and QM counterparts. Blue and yellow dots represent the train and test set, respectively. The color bar represents the density of the data points in the training set.

= 0.043 eV, R = 0.96) on test set are achieved, while the prediction on PLQY is less precise but still acceptable. This is aroused from the fact that the former three are pure electronic structure properties that only depend on electronic structure calculation, while PLQY is obtained by calculating the radiative and nonradiative decay rate constants via the rate formalizm (as detailed in the Supporting Information). These two quantities are closely related to both electronic structure and photophysical properties of the light-emitting state and are intrinsically difficult to be accurately computed and predicted. We verify the robustness of Uni-Mol by performing four independent training to predict PLQY. As shown in Table S2 (Supporting Information), the MAEs of PLQY given

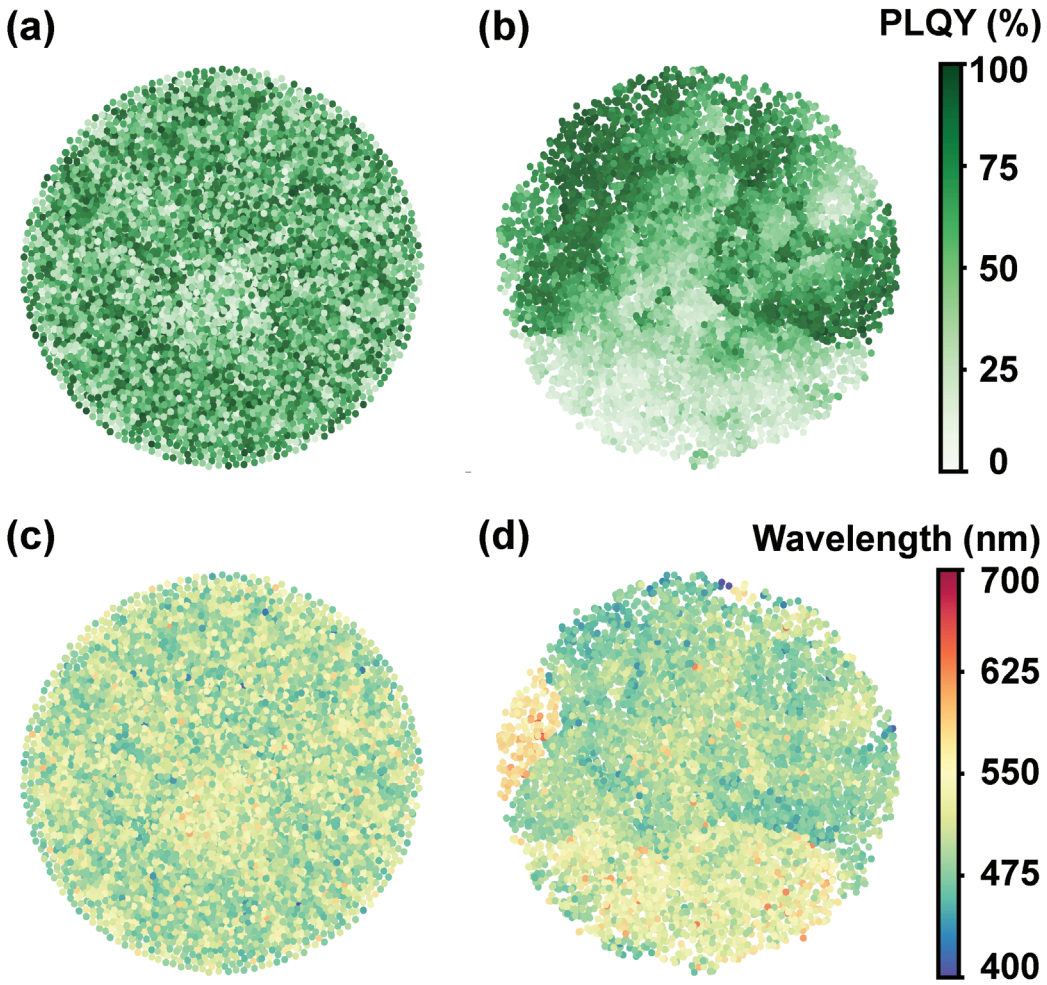

Figure 4. Distribution of PLQY (a,b) and emission wavelength (c,d) in 2D projections of Uni-Mol embedding layer via t-SNE. Data points in (a) and (c) are predicted via the Uni-Mol model with random weight, while those in (b) and (d) are predicted via trained Uni-Mol model.

by these four models are basically lower than 10%, with a small standard deviation of 1%. It can be seen that the overall MAE of PLQY given by Uni-Mol (9.4%) reaches a desired accuracy for high throughput screening.

To further justify the strength of Uni-Mol in predicting the optical properties of OLED molecules, we employ another two ML models, namely, the Deep Potential (DP)[73] and E3NN,[40] on our Ir(III) complexes dataset. For DP and E3NN, we randomly split the dataset to training and test sets and the ratio of training and test set is set to 9:1. Details of hyperparameters of DP and E3NN are provided in the Supporting Information. The mean absolute errors (MAEs) of DP, E3NN, and Uni-Mol in predicting the four QM properties are listed in Table S3 (Supporting Information). For all investigated properties in our dataset, the DP shows the worst accuracy, which is because DP only employs pair distances and angles of molecules to describe the spatial information. The E3NN employs the E(3)-equivariant convolutions and message passing algorithm, resulting in a more information-rich and reliable representation for the spatial information.[40] The accuracy of E3NN is therefore remarkably higher than that of DP. Benefitting from the pre-training process, the accuracy of Uni-Mol on four QM properties are further improved compared with that of E3NN as shown in Table S3 (Supporting Information). Table S3 (Supporting Information) also illustrates the poor accuracy of Uni-Mol without pre-training, which is merely on par with DP, thereby signifying the necessity of pre-training in the Uni-Mol training process. Altogether, these results evince that our pre-trained and fine-tuned Uni-Mol model can be applied as a powerful tool to model the structure-property relationship for optical molecules with limited QM data.

After training the Uni-Mol model, we apply t-SNE, one of the best techniques for high-dimensional data reduction and visualization,[74] to project the embedding layer (512 dimensions) of the Uni-Mol model into two dimensions. **Figure 4** shows the distribution of 8500 randomly selected candidates in the 2D space, where the upper and lower color bars represent PLQY and emission wavelength, respectively. Figure 4a,c display the t-SNE projected results of PLQY and wavelength before training, which are purely randomly distributed, while data points after training in Figure 4b,d are well-clustered into distinct regions with respect to the value of PLQY or emission wavelength, elucidating the existence of the structure-property relationship established by the Uni-Mol model. It can be seen that data points with PLQY greater than 70% are concentrated in an area of about 30% of

Table 2. The adiabatic excitation energy ranges that are applied to filter out four different colors.

Color	Red	Yellow	Green	Blue
E_{ad} min. (eV)	2.00	2.32	2.55	2.84
E_{ad} max. (eV)	2.21	2.40	2.63	3.27

the whole 2D plot, and data points with varied wavelengths distribute in different regions. In addition, it is also found that the deep green area (PLQY larger than 70%) in Figure 4b almost coincides with the blue/green area (wavelength shorter than 500 nm) in Figure 4d, while the region of red/orange emitters are basically located at the low PLQY region. This is consistent with the experimental observation that OLED materials with short emission wavelength usually have high PLQY. Figure 4d also shows that the emission wavelengths of most selected candidate molecules are within 450–650 nm range, while the pure red or pure blue emitters are relatively rare. In fact, only a tiny proportion of the candidate pool would meet the specific screening criteria for display purpose, which will be introduced in next section.

3.3. Screen-Out of Novel Emitters

To search for potentially outstanding emitters, we apply Uni-Mol model to our candidate pool and perform the screening according to four screening criteria, i.e., PLQY, E_{ad}, the number of ligand types, and the number of heavy atoms. First, we screen out red, yellow, green, and blue emitters based on the Uni-Mol predicted emission energy of all 1.6 million candidates. Note that the experimental counterpart of the theoretically computed E_{ad} is close to the crossing point of the absorption/fluorescence spectra, of which the wavelength is shorter compared to the maximum emission wavelength. Therefore, as shown in **Table 2**, E_{ad} ranges that applied to screen out four different colors are set a bit higher compared to the normal emission energy of corresponding emitters. Based on these energy ranges, 9772, 90664, 384170, and 61329 candidates are singled out, corresponding to red, yellow, green, and blue emitters, respectively. Next, molecules with PLQY larger than 40% are selected from the red and yellow set, and molecules with PLQY larger than 80% are selected from the green and blue set, ensuring a favored high quantum efficiency. From a practical point of view, candidates that are difficult to be synthesized and/or further processed are eliminated, such as those with three different types of ligands and/or with unreasonably large molecular weights. The screened-out candidates according to all aforementioned criteria are summarized in the Supporting Information. It can be seen that our search is able to recover some well-known phosphorescent emitters, such as $Ir(Fppy)_2(acac)$, FIrpic[75] and a few $(dfppy)_2Ir(NHC)$ complexes,[70] which essentially validates the rationality of our theoretical protocol.

Molecular structures of the most promising candidates for both display and lighting purposes, which have not been previously reported, are shown in **Figure 5**, with the theoretically predicted spectra and CIE coordinates. Three candidates with red, green, and blue emission are predicted with high performance for display as shown in Figure 5a,b. For display purpose, narrower full-width at half-maximum (FWHM) of the emitter, especially for green and blue emitters, is desired to cover a wider color gamut and provide more vivid hue. Based on our screening procedure, the green emitter **89_89_228** and the blue emitter **152_189_189** stand out, with remarkably narrower FWHM (38 and 52 nm, respectively) compared to common green and blue phosphorescent emitters. The CIE coordinates of these two emitters are (0.22, 0.66) and (0.15, 0.05), respectively. The screened out red emitter **41_41_75**, though with slightly broader FWHM, has the main emission peak at 611 nm, which corresponds to a satisfactory CIE coordinate (0.66, 0.34) for red emitters. It can be seen

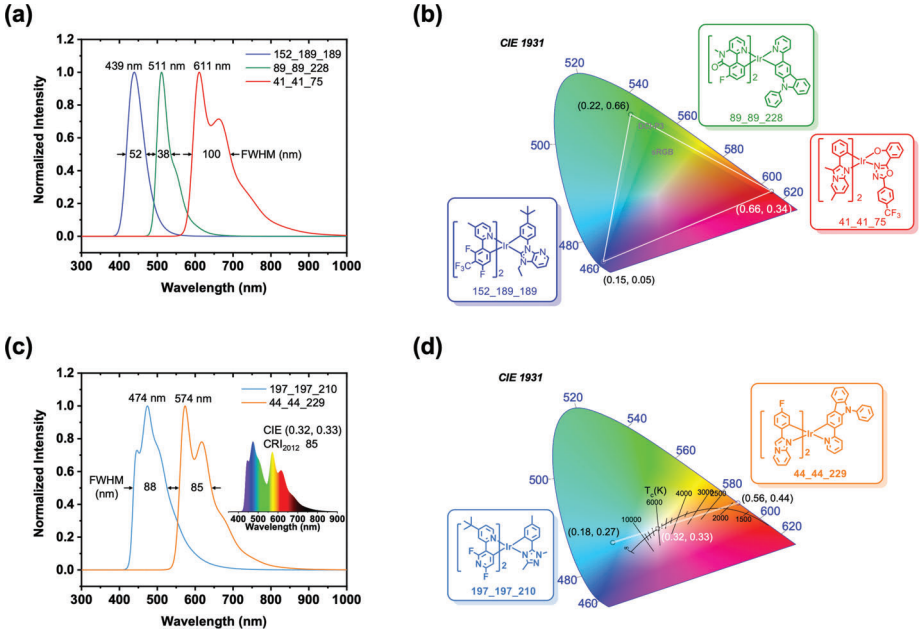

Figure 5. Theoretical predicted emission spectra and CIE coordinates of screened out candidates for both display (a,b) and lighting (c,d) purposes.

in Figure 5b that the color gamut covered by these three emitters is close to that defined by DCI-P3 and significantly wider than that defined by sRGB.[76]

Contrary to displays, white lighting devices require broader emission spectra to achieve higher lighting quality, which could be quantified by color rendering index (CRI). Dual-color based on blue and yellow light is widely applied as a low-cost white lighting protocol. According to our HTVS workflow, two candidates are screened out for lighting purpose, i.e., the sky blue emitter **197_197_210** and the yellow-orange emitter **44_44_229**. As shown in Figure 5c, the predicted emission spectra of these two molecules exhibit main peaks at 474 and 574 nm, respectively, both of which acquire broad FWHM larger than 85 nm. The corresponding CIE coordinates of these two candidates read (0.18, 0.27) and (0.56, 0.44), respectively, as depicted in Figure 5d. By tuning the mixing proportion of these two colors, pure white emission with CIE coordinate (0.32, 0.33) can be achieved as shown in the inset of Figure 5c,d. The corresponding CRI2012 index [77] is calculated as 85, which could be attractive for daily light source with balanced quality and cost. Based on our theoretical analysis, we foresee good performance of these selected candidates that are yet to be experimentally synthesized and explored for both display and lighting.

4. Conclusion

In this work, we have applied Uni-Mol, an advanced MRL algorithm, combined with accurately benchmarked QM calculations to perform HTVS over millions of Ir(III) complex emitters. The whole screening process consists of three automatized workflows, i) the generation of tremendous candidates with Ir atom and a random combination of 278 reported ligands, ii) the generation of 3D inputs and labels for Uni-Mol, which is accomplished via the semi-empirical method on all candidates with rational structures and the high-level QM calculation over around 1500 molecules, and iii) the pre-training and fine-tuning process of Uni-Mol. Based on this cascade of automatic workflows, we are able to screen out a decent amount of promising Ir(III) complexes for both display and lighting purposes, which are worth further experimental investigation. To the best of our knowledge, this is the first work that applies MRL to organic luminescent materials, and our screening protocol can be effortlessly transferred to other organic materials, impulsing the design of novel materials that are otherwise shielded by conventional mindset and/or limited computational and experimental resources.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The authors gratefully acknowledge the funding support from the AI for Science Institute, Beijing (AISI) and DP Technology Corporation. The computing resources for this work were provided by the Bohrium Cloud Platform (https://bohrium.dp.tech, which was supported by DP Technology), the Hefei Advanced Computing Center of Sugon, and the High-Performance Computing Platform at Peking University.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Z.C. and J.L. contributed equally to this work. Q.O. and Z.Z. conceived the project. Z.C. and T.J. carried out the construction of the candidate pool and quantum chemistry calculations under the assistance of M.C. and F.D. J.L. carried out the Uni-Mol pre-training and fine-tuning process under the assistance of Z.G. and G.K.. All authors contributed to data analysis and writing the paper.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

high-throughput virtual screening, OLED, quantum mechanical calculation, representation learning

Received: May 31, 2023
Published online: July 25, 2023

[1] C. W. Tang, S. A. VanSlyke, *Appl. Phys. Lett.* **1987**, *51*, 913.
[2] T. Kusamoto, H. Nishihara, *Nature* **2018**, *563*, 480.
[3] X. Li, J. Zhang, Z. Zhao, L. Wang, H. Yang, Q. Chang, N. Jiang, Z. Liu, Z. Bian, W. Liu, Z. Lu, C. Huang, *Adv. Mater.* **2018**, *30*, 1705005.
[4] T. Hatakeyama, K. Shiren, K. Nakajima, S. Nomura, S. Nakatsuka, K. Kinoshita, J. Ni, Y. Ono, T. Ikuta, *Adv. Mater.* **2016**, *28*, 2777.
[5] L. Wang, Z. Zhao, G. Zhan, H. Fang, H. Yang, T. Huang, Y. Zhang, N. Jiang, L. Duan, Z. Liu, Z. Bian, Z. Lu, C. Huang, *Light Sci. Appl.* **2020**, *9*, 157.
[6] H. J. Kim, T. Yasuda, *Adv. Opt. Mater.* **2022**, *10*, 2201714.
[7] G. Hong, X. Gan, C. Leonhardt, Z. Zhang, J. Seibert, J. M. Busch, S. Bräse, *Adv. Mater.* **2021**, *33*, 2005630.
[8] S. Curtarolo, G. L. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, O. Levy, *Nat. Mater.* **2013**, *12*, 191.
[9] E. O. Pyzer-Knapp, J. W. Pitera, P. W. Staar, S. Takeda, T. Laino, D. P. Sanders, J. Sexton, J. R. Smith, A. Curioni, *npj Comput. Mater.* **2022**, *8*, 84.
[10] S. Axelrod, D. Schwalbe-Koda, S. Mohapatra, J. Damewood, K. P. Greenman, R. Gómez-Bombarelli, *Acc. Mater. Res.* **2022**, *3*, 343.
[11] N. C. Forero-Martinez, K.-H. Lin, K. Kremer, D. Andrienko, *Adv. Sci.* **2022**, *9*, 2200825.
[12] X. Zhu, M. Ge, T. Sun, X. Yuan, Y. Li, *J. Phys. Chem. Lett.* **2023**, *14*, 2215.
[13] Y. Liu, X. Tan, J. Liang, H. Han, P. Xiang, W. Yan, *Adv. Funct. Mater.* **2023**, *33*, 2214271.
[14] P. Gorai, D. Krasikov, S. Grover, G. Xiong, W. K. Metzger, V. Stevanović, *Sci. Adv.* **2023**, *9*, eade3761.
[15] M. Aykol, S. Kim, V. I. Hegde, D. Snydacker, Z. Lu, S. Hao, S. Kirklin, D. Morgan, C. Wolverton, *Nat. Commun.* **2016**, *7*, 13779.
[16] S. Sun, N. T. Hartono, Z. D. Ren, F. Oviedo, A. M. Buscemi, M. Layurova, D. X. Chen, T. Ogunfunmi, J. Thapa, S. Ramasamy, C. Settens, B. L. DeCost, A. G. Kusne, Z. Liu, S. I. P. Tian, I. M. Peters, J.-P. Correa-Baena, T. Buonassisi, *Joule* **2019**, *3*, 1437.
[17] A. Benayad, D. Diddens, A. Heuer, A. N. Krishnamoorthy, M. Maiti, F. L. Cras, M. Legallais, F. Rahmanian, Y. Shin, H. Stein, M. Winter, C. Wölke, P. Yan, I. Cekic-Laskovic, *Adv. Energy Mater.* **2022**, *12*, 2102678.

[18] X. Zhang, Z. Zhang, D. Wu, X. Zhang, X. Zhao, Z. Zhou, *Small Methods* **2018**, *2*, 1700359.
[19] S. Lu, Q. Zhou, Y. Guo, Y. Zhang, Y. Wu, J. Wang, *Adv. Mater.* **2020**, *32*, 2002658.
[20] N. Mounet, M. Gibertini, P. Schwaller, D. Campi, A. Merkys, A. Marrazzo, T. Sohier, I. E. Castelli, A. Cepellotti, G. Pizzi, N. Marzari, *Nat. Nanotechnol.* **2018**, *13*, 246.
[21] G. L. Hart, T. Mueller, C. Toher, S. Curtarolo, *Nat. Rev. Mater.* **2021**, *6*, 730.
[22] R. Feng, C. Zhang, M. C. Gao, Z. Pei, F. Zhang, Y. Chen, D. Ma, K. An, J. D. Poplawsky, L. Ouyang, Y. Ren, J. A. Hawk, M. Widom, P. K. Liaw, *Nat. Commun.* **2021**, *12*, 4329.
[23] Y. Li, J. Yang, R. Zhao, Y. Zhang, X. Wang, X. He, Y. Fu, L. Zhang, *J. Am. Chem. Soc.* **2022**, *144*, 16656.
[24] Y. Gan, N. Miao, P. Lan, J. Zhou, S. R. Elliott, Z. Sun, *J. Am. Chem. Soc.* **2022**, *144*, 5878.
[25] C. Ling, Y. Cui, S. Lu, X. Bai, J. Wang, *Chem* **2022**, *8*, 1575.
[26] S. Wang, H. Gao, L. Li, K. San Hui, D. A. Dinh, S. Wu, S. Kumar, F. Chen, Z. Shao, K. N. Hui, *Nano Energy* **2022**, *100*, 107517.
[27] J. Li, P. Du, S. Li, J. Liu, M. Zhu, Z. Tan, M. Hu, J. Luo, D. Guo, L. Ma, et al., *Adv. Funct. Mater.* **2019**, *29*, 1903607.
[28] R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, T. D. Hirzel, D. Duvenaud, D. Maclaurin, M. A. Blood-Forsythe, H. S. Chae, M. Einzinger, D.-G. Ha, T. Wu, G. Markopoulos, S. Jeon, H. Kang, H. Miyazaki, M. Numata, S. Kim, W. Huang, S. I. Hong, M. Baldo, R. P. Adams, A. Aspuru-Guzik, *Nat. Mater.* **2016**, *15*, 1120.
[29] W. Sun, Y. Zheng, K. Yang, Q. Zhang, A. A. Shah, Z. Wu, Y. Sun, L. Feng, D. Chen, Z. Xiao, S. Lu, Y. Li, K. Sun, *Sci. Adv.* **2019**, *5*, eaay4275.
[30] R. Xia, C. J. Brabec, H.-L. Yip, Y. Cao, *Joule* **2019**, *3*, 2241.
[31] J. Westermayr, J. Gilkes, R. Barrett, R. J. Maurer, *Nat. Comput. Sci.* **2023**, *3*, 139.
[32] D. Behrendt, S. Banerjee, C. Clark, A. M. Rappe, *J. Am. Chem. Soc.* **2023**, *145*, 4730.
[33] F. Tang, S. Ono, X. Wan, H. Watanabe, *Phys. Rev. Lett.* **2022**, *129*, 027001.
[34] Y. Chen, X. Bai, D. Liu, X. Fu, Q. Yang, *ACS Appl. Mater. Interfaces* **2022**, *14*, 24980.
[35] H. Lischka, D. Nachtigallova, A. J. Aquino, P. G. Szalay, F. Plasser, F. B. Machado, M. Barbatti, *Chem. Rev.* **2018**, *118*, 7293.
[36] S. Axelrod, E. Shakhnovich, R. Gómez-Bombarelli, *Nat. Commun.* **2022**, *13*, 3440.
[37] J. Westermayr, P. Marquetand, *Chem. Rev.* **2020**, *121*, 9873.
[38] M. Karelson, V. S. Lobanov, A. R. Katritzky, *Chem. Rev.* **1996**, *96*, 1027.
[39] T. Xie, J. C. Grossman, *Phys. Rev. Lett.* **2018**, *120*, 145301.
[40] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, B. Kozinsky, *Nat. Commun.* **2022**, *13*, 2453.
[41] H. Ren, Q. Zhang, Z. Wang, G. Zhang, H. Liu, W. Guo, S. Mukamel, J. Jiang, *PNAS* **2022**, *119*, e2202713119.
[42] J. S. Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev, A. E. Roitberg, *Nat. Commun.* **2019**, *10*, 2903.
[43] G. Pesciullesi, P. Schwaller, T. Laino, J.-L. Reymond, *Nat. Commun.* **2020**, *11*, 4874.
[44] Y. Rong, Y. Bian, T. Xu, W. Xie, Y. Wei, W. Huang, J. Huang, *Adv. Neural Inf. Process* **2020**, *33*, 12559.
[45] Y. Wang, J. Wang, Z. Cao, A. Barati Farimani, *Nat. Mach. Intell.* **2022**, *4*, 279.
[46] J. Devlin, M.-W. C. Kenton, L. K. Toutanova, in *Proceedings of NAACL-HLT. Association for Computational Linguistics*, Stroudsburg, Pennsylvania **2019**.
[47] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, *OpenAI Blog* **2019**, *1*, 9.
[48] C. Ying, T. Cai, S. Luo, S. Zheng, G. Ke, D. He, Y. Shen, T.-Y. Liu, *Adv. Neural Inf. Process* **2021**, *34*, 28877.
[49] G. Zhou, Z. Gao, Q. Ding, H. Zheng, H. Xu, Z. Wei, L. Zhang, G. Ke, in *The Eleventh International Conference on Learning Representations*, ICLR, San Juan, Puerto Rico **2023**.
[50] T.-Y. Li, J. Wu, Z.-G. Wu, Y.-X. Zheng, J.-L. Zuo, Y. Pan, *Coord. Chem. Rev.* **2018**, *374*, 55.
[51] L. Turcani, A. Tarzia, F. T. Szczypiński, K. E. Jelfs, *J. Chem. Phys.* **2021**, *154*, 214102.
[52] C. Bannwarth, S. Ehlert, S. Grimme, *J. Chem. Theory Comput.* **2019**, *15*, 1652.
[53] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, et al., Gaussian 16 Revision C.01, Gaussian Inc. Wallingford CT, **2016**.
[54] F. Neese, *WIREs Comput. Mol. Sci.* **2012**, *2*, 73.
[55] F. Neese, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, e1327.
[56] Z. Shuai, Q. Peng, *Natl. Sci. Rev.* **2017**, *4*, 224.
[57] Q. Peng, Y. Yi, Z. Shuai, J. Shao, *J. Am. Chem. Soc.* **2007**, *129*, 9333.
[58] Z. Shuai, *Chin. J. Chem.* **2020**, *38*, 1223.
[59] S.-J. Yoon, J. W. Chung, J. Gierschner, K. S. Kim, M.-G. Choi, D. Kim, S. Y. Park, *J. Am. Chem. Soc.* **2010**, *132*, 13675, pMID: 20839795.
[60] Z. Wu, L. Ma, P. Liu, C. Zhou, S. Ning, A. El-Shafei, X. Zhao, X. Hou, *J. Phys. Chem A* **2013**, *117*, 10903.
[61] L. Ma, Y. Yu, B. Jiao, X. Hou, Z. Wu, *Phys. Chem. Chem. Phys.* **2018**, *20*, 19515.
[62] K. Chao, K. Shao, T. Peng, D. Zhu, Y. Wang, Y. Liu, Z. Su, M. R. Bryce, *Journal of Materials Chemistry C* **2013**, *1*, 6800.
[63] R. Wang, D. Liu, H. Ren, T. Zhang, X. Wang, J. Li, *J. Mater. Chem.* **2011**, *21*, 15494.
[64] X. Ma, J. Liang, F. Bai, K. Ye, J. Xu, D. Zhu, M. R. Bryce, *Eur. J. Inorg. Chem.* **2018**, *42*, 4614.
[65] J. Lee, H. Park, K.-M. Park, J. Kim, J.-Y. Lee, Y. Kang, *Dyes Pigm.* **2015**, *123*, 235.
[66] J. Zhao, Y. Yu, X. Yang, X. Yan, H. Zhang, X. Xu, G. Zhou, Z. Wu, Y. Ren, W.-Y. Wong, *ACS Appl. Mater.Interfaces* **2015**, *7*, 24703.
[67] Y.-M. Jing, Y. Zheng, *New J. Chem.* **2017**, *41*, 3029.
[68] J.-H. Jou, Y.-X. Lin, S.-H. Peng, C.-J. Li, Y.-M. Yang, C.-L. Chin, J.-J. Shyue, S.-S. Sun, M. Lee, C.-T. Chen, M.-C. Liu, C.-C. Chen, G.-Y. Chen, J.-H. Wu, C.-H. Li, C.-F. Sung, M.-J. Lee, J.-P. Hu, *Adv. Funct. Mater.* **2014**, *24*, 555.
[69] G. Tan, S. Chen, N. Sun, Y. Li, D. Fortin, W.-Y. Wong, H.-S. Kwok, D. Ma, H. Wu, L. Wang, P. D. Harvey, *J. Mater. Chem. C* **2013**, *1*, 808.
[70] T.-Y. Li, X. Liang, L. Zhou, C. Wu, S. Zhang, X. Liu, G.-Z. Lu, L.-S. Xue, Y.-X. Zheng, J.-L. Zuo, *Inorg. Chem.* **2015**, *54*, 161.
[71] J. F. Joung, M. Han, M. Jeong, S. Park, *Sci. Data* **2020**, *7*, 295.
[72] J. F. Joung, M. Han, J. Hwang, M. Jeong, D. H. Choi, S. Park, *JACS Au* **2021**, *1*, 427.
[73] L. Zhang, J. Han, H. Wang, R. Car, E. Weinan, *Phys. Rev. Lett.* **2018**, *120*, 143001.
[74] L. Van der Maaten, G. Hinton, *J. Mach. Learn Res.* **2008**, *9*, 11.
[75] C. Adachi, R. C. Kwong, P. Djurovich, V. Adamovich, M. A. Baldo, M. E. Thompson, S. R. Forrest, *Appl. Phys. Lett.* **2001**, *79*, 2082.
[76] B. Ryu, K. Kim, Y. Ha, J. Bae, S. Lee, J. Song, K. Lee, J. Lee, K. Kim, H. Kim, *J. Inf. Disp.* **2014**, *15*, 65.
[77] K. Smet, J. Schanda, L. Whitehead, R. Luo, *Light. Res. Technol.* **2013**, *45*, 689.