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[*,‡]

†*School of Mathematical Sciences, Peking University, Beijing 100871 China*

‡*AI for Science Institute, Beijing 100084, P.R. China*

¶*School of Advanced Energy, Sun Yat-Sen University, Shenzhen 518107, China*

§*MOE Key Laboratory of Organic OptoElectronics and Molecular Engineering, Department of Chemistry, Tsinghua University, Beijing 100084, China.*

‖*HEDPS, CAPT, College of Engineering and School of Physics, Peking University, Beijing 100871, P.R. China*

⊥*DP Technology, Beijing 100080, P.R. China*

#*Contributed equally to this work*

E-mail: zhaozf@bjaisi.com; ouq@bjaisi.com

Abstract

Organic light-emitting diodes (OLEDs) have gained widespread commercial use, yet there is a continuous need to identify innovative emitters that offer higher efficiency and broader color gamut. To effectively screen out promising OLED molecules that are yet to be synthesized, we perform a representation learning aided high throughput virtual screening (HTVS) over millions of Ir(III) complexes, a prototypical type of phosphorescent OLED material, constructed via a random combination of 278 reported ligands.

We successfully screen 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 our 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-tune 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.

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 rely on expertised 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 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.

One of the machine learning (ML) algorithms that successfully tackles this issue for molecular systems is called molecular representation learning (MRL), or pre-training, self-supervised learning.[38–42] In such algorithm, the property-predicting model is constructed with a pre-training learning process of tremendous unlabeled data, followed by the application of fine-tuning schema with labeled data on multiple downstream tasks. Proposed in 2022, Uni-Mol is an advanced MRL scheme that directly takes the 3D structures (instead of the 2D or 1D representation) of interested molecules as input, making it a competitive tool in predicting structural-related properties of various materials.[43] In fact, by taking advantage of enormous 3D conformations of the target systems, Uni-Mol rivals the performances of the state-of-the-art (SOTA) methods in predicting energetic properties of open-source organic and medical molecules.[43]

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,

enabling an efficient means of material design in a variety of fields.

Theory and computational details

The overall workflow of our HTVS process is schematized in Figure 1. With 278 bidentate ligands collected from previously published experimental studies on Ir(III) complex,[44] millions of candidate molecules are constructed using stk[45] and further optimized via density functional tight binding approach (GFN2-xTB).[46] A structure rationality filter is then introduced to filter out unphysical structures based on atomic distances and/or ligand angles. As such, around 1.6 million candidates are constructed, of which the coordinates and symbols are served as Uni-Mol inputs as well as the initial guesses for QM calculations, while the energies of highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) at GFN2-xTB level are served as labels for Uni-Mol pre-training task.

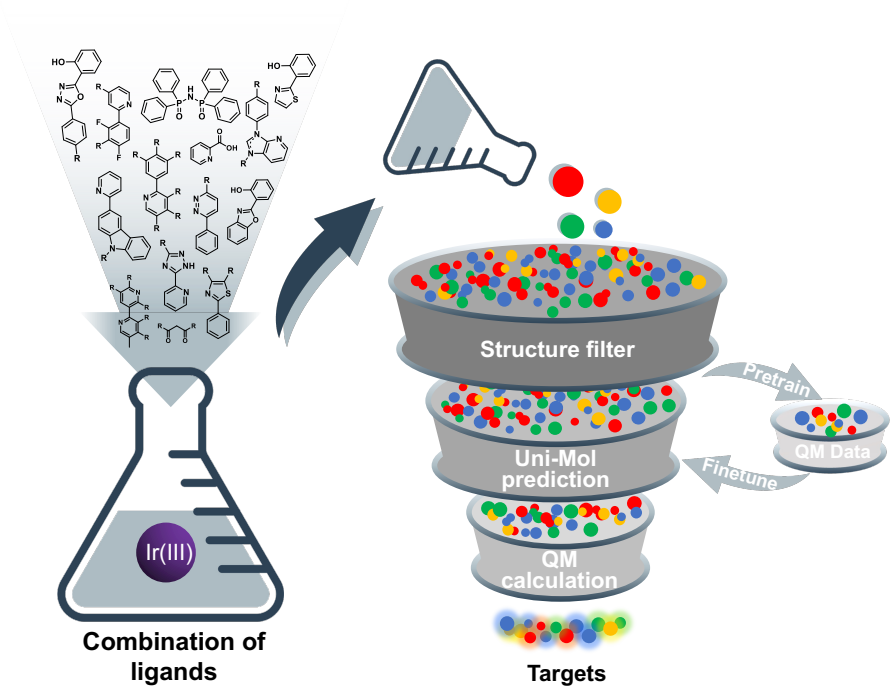


Figure 1: Schematic HTVS workflow for Ir(III) complex emitters.

Due to the high computational cost of QM calculations, they are 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 can be seen in Figure S2. Four QM predicted properties, HOMO, LUMO, the adiabatic excitation energy of T_1, and the photoluminescence quantum yield (PLQY), are utilized to fine-tune the pre-trained model. All electronic structure calculations are performed with quantum chemistry packages Gaussian16[47] and ORCA,[48,49] while the photophysical properties including emission spectra and various rate constants are calculated via the thermal vibration correlation function (TVCF) method in the molecular material property prediction package MOMAP.[50–52] The accuracy of our QM calculation process is 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 are provided in Figure S3.

After the pre-training and fine-tuning process, the established Uni-Mol model is applied to single out potentially outstanding candidates, followed by further QM validation on these molecules. Target molecules for displaying and lighting purposes are then screened out with corresponding criteria (as detailed in the next section). Additional details of our HTVS workflow (Figure S1) and the applied electronic structure theory are provided in the Supporting Information.

Results and discussion

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,[62] 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

Table 1: PLQY and adiabatic excitation energy of ten OLED molecules predicted by QM calculation and the corresponding experimental values.

Molecule ID	PLQY		E_{ad} (eV)	
	QM	Exp.	QM	Exp.[a]
$(ppy)_2Ir(oz)$[53]	0.30	0.55	2.60	2.58
$Ir(dpt)_3$[3]	0.99	0.64	3.14	3.13
Complex 2[54]	0.39	0.35	2.62	2.44
$(mdp)_2Ir(acac)$[55]	0.76	0.85	2.40	2.37
Complex 1[56]	0.78	0.78	2.97	3.12
IrS-5F[57]	0.67	0.95	2.47	2.35
Ir1[58]	0.99	0.93	2.46	2.75
2FBNO[59]	0.83	0.71	2.37	2.45
Complex 2[60]	0.90	0.73	2.63	2.94
Ir5b[61]	0.99	0.69	2.91	2.96

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

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.[63] 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.

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 GFN-xTB level as labels. By dividing the dataset into training, validation, and test sets with the ratio of 8:1:1, the R for HOMO and LUMO on the test set is 0.995 and 0.934, respectively, which evinces the consistency between Uni-Mol prediction and GFN-xTB calculations. Such pre-training process is followed by the fine-

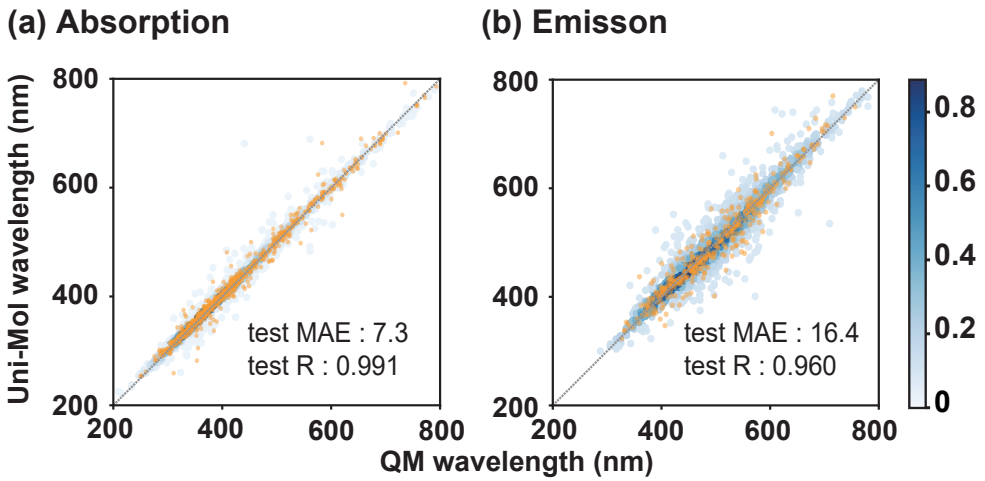


Figure 2: Correlation between the Uni-Mol predicted values [(a) absorption and (b) emission wavelength] and the experimental values on an open-source database. The blue and yellow dots represent the train and test set, respectively. The color bar represents the density of data points in the absorption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

tuning with four QM predicted properties, i.e., HOMO, LUMO, adiabatic excitation energy of T_1, and PLQY, and 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 accurate predictions of HOMO (MAE = 0.067 eV, R = 0.95), LUMO (MAE = 0.114 eV, R = 0.87), and E_{ad} (MAE = 0.043 eV, R = 0.96) on test set are achieved, while the prediction on PLQY is less satisfactory. 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 rate formalism. 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. That being said, the MAE and R for PLQY are 9.4% and 0.87, respectively, which are still acceptable for high throughput screening.

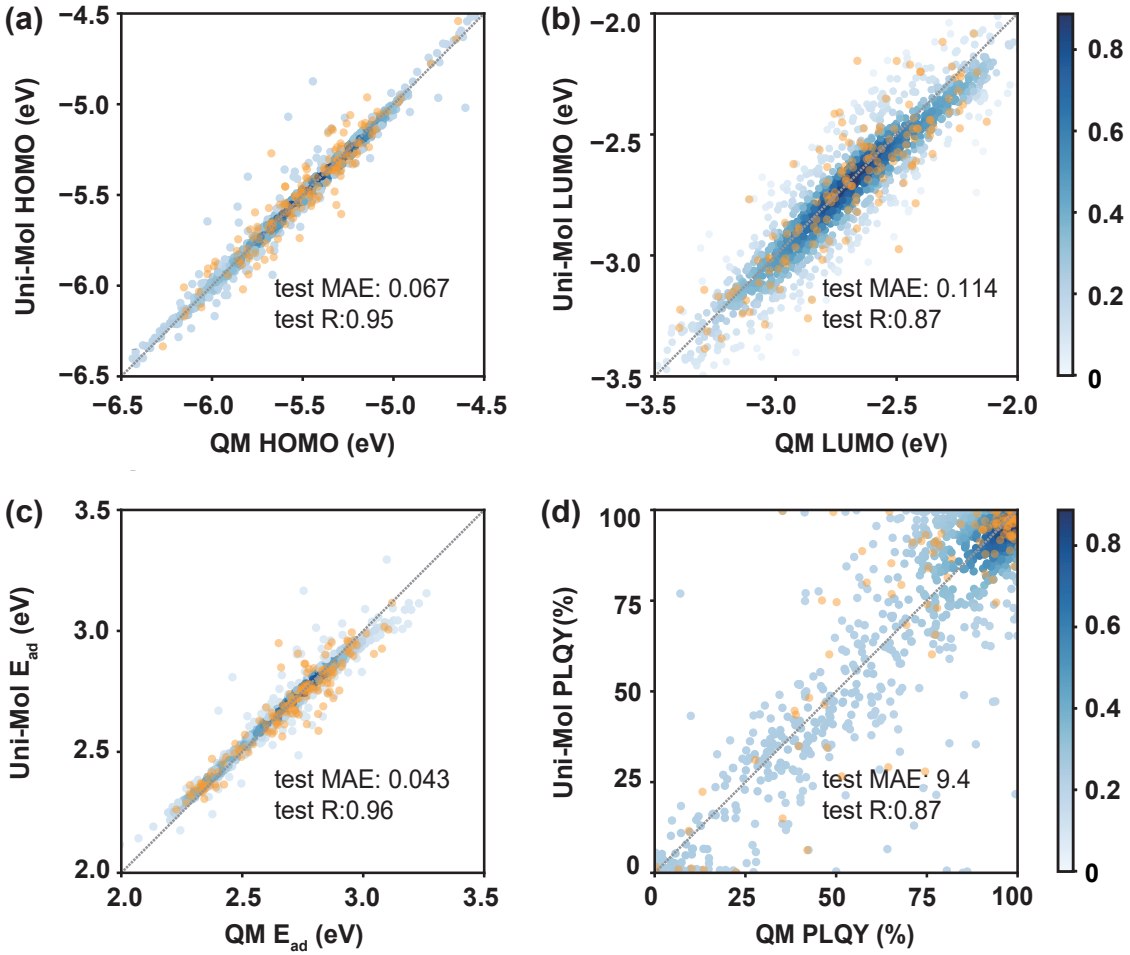


Figure 3: Correlation between the Uni-Mol predicted values [(a) HOMO, (b) LUMO, (c) adiabatic excitation energy, and (d) PLQY] and QM counterparts. The blue and yellow dots represent the train and test set, respectively. The color bar represents the density of data points in the HOMO, LUMO, adiabatic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

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 $\mathrm{Ir(Fppy)_2(acac)}$, FIrpic[64] and a few $\mathrm{(dfppy)_2Ir(NHC)}$ complexes,[61] which essentially validates the rationality of our theoretical protocol.

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

Molecular structures of the most promising candidates for both display and lighting purposes, which have not been previously reported, are shown in Figure 4, 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 4(a) and (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 nm 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 in Figure 4(b) 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.[65]

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 4(c), the predicted emission spectra of these two molecules exhibit main peaks at 474 nm 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. 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 4(c). The corresponding CRI2012 index[66] 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.

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

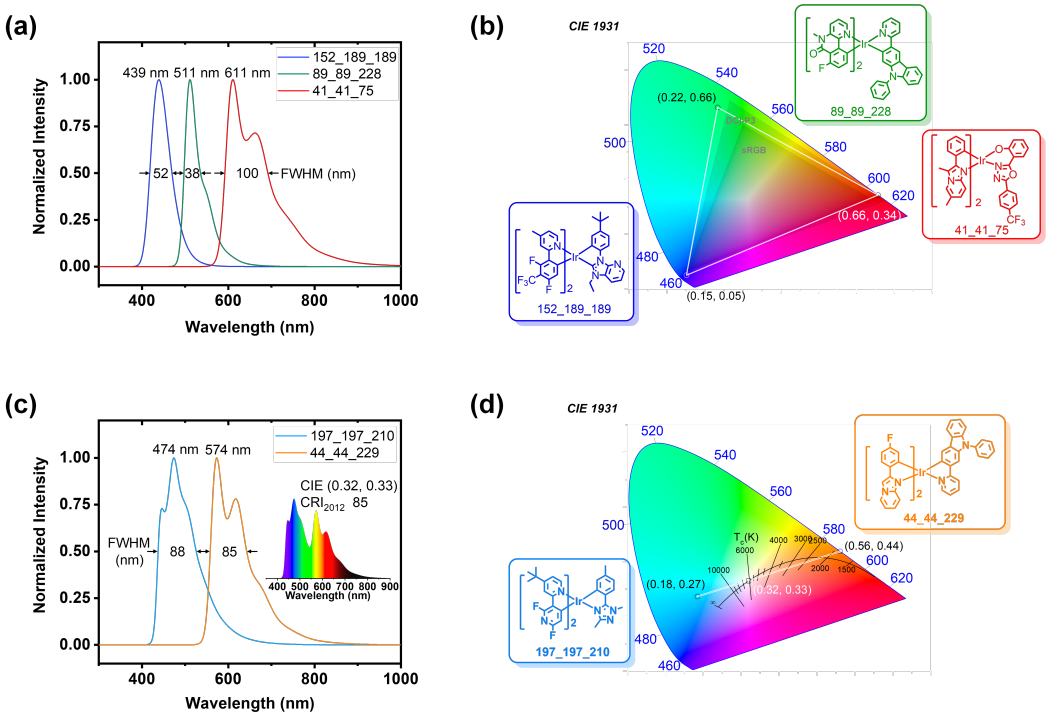


Figure 4: Theoretical predicted emission spectra and CIE coordinates of screened out candidates for both display [(a) and (b)] and lighting [(c) and (d)] purposes.

the semi-empirical method on all candidates with rational structures and the high-level QM calculation over around 1500 molecules, and (iii) the pre-train and fine-tune 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.

Acknowledgement

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

Supporting Information Available

Detailed workflow of the HTVS procedure, applied semi-empirical and *ab initio* methodologies, SMILES of 278 reported ligands, 3D information of candidates with red, yellow, green and blue colors, heavy atom distribution of the whole candidate pool, and 3D coordinates of screened-out candidates shown are provided in the Supporting Information.

References

(1) Tang, C. W.; VanSlyke, S. A. Organic Electroluminescent Diodes. *Applied physics letters* **1987**, *51*, 913–915.

(2) Kusamoto, T.; Nishihara, H. Efficiency Breakthrough for Radical LEDs. *Nature* **2018**, *563*, 480–481.

(3) Li, X.; Zhang, J.; Zhao, Z.; Wang, L.; Yang, H.; Chang, Q.; Jiang, N.; Liu, Z.; Bian, Z.; Liu, W., et al. Deep Blue Phosphorescent Organic Light-Emitting Diodes with CIEy Value of 0.11 and External Quantum Efficiency up to 22.5%. *Advanced Materials* **2018**, *30*, 1705005.

(4) Hatakeyama, T.; Shiren, K.; Nakajima, K.; Nomura, S.; Nakatsuka, S.; Kinoshita, K.; Ni, J.; Ono, Y.; Ikuta, T. Ultrapure Blue Thermally Activated Delayed Fluorescence

Molecules: Efficient HOMO–LUMO Separation by the Multiple Resonance Effect. *Advanced Materials* **2016**, *28*, 2777–2781.

(5) Wang, L.; Zhao, Z.; Zhan, G.; Fang, H.; Yang, H.; Huang, T.; Zhang, Y.; Jiang, N.; Duan, L.; Liu, Z., et al. Deep-Blue Organic Light-Emitting Diodes Based on a Doublet d–f Transition Cerium (III) Complex with 100% Exciton Utilization Efficiency. *Light: Science & Applications* **2020**, *9*, 157.

(6) Kim, H. J.; Yasuda, T. Narrowband Emissive Thermally Activated Delayed Fluorescence Materials. *Advanced Optical Materials* **2022**, *10*, 2201714.

(7) Hong, G.; Gan, X.; Leonhardt, C.; Zhang, Z.; Seibert, J.; Busch, J. M.; Bräse, S. A Brief History of OLEDs—Emitter Development and Industry Milestones. *Advanced Materials* **2021**, *33*, 2005630.

(8) Curtarolo, S.; Hart, G. L.; Nardelli, M. B.; Mingo, N.; Sanvito, S.; Levy, O. The High-Throughput highway to Computational Materials Design. *Nature Materials* **2013**, *12*, 191–201.

(9) Pyzer-Knapp, E. O.; Pitera, J. W.; Staar, P. W.; Takeda, S.; Laino, T.; Sanders, D. P.; Sexton, J.; Smith, J. R.; Curioni, A. Accelerating Materials Discovery Using Artificial Intelligence, High Performance Computing and Robotics. *npj Computational Materials* **2022**, *8*, 84.

(10) Axelrod, S.; Schwalbe-Koda, D.; Mohapatra, S.; Damewood, J.; Greenman, K. P.; Gómez-Bombarelli, R. Learning matter: Materials design with machine learning and atomistic simulations. *Accounts of Materials Research* **2022**, *3*, 343–357.

(11) Forero-Martinez, N. C.; Lin, K.-H.; Kremer, K.; Andrienko, D. Virtual screening for organic solar cells and light emitting diodes. *Advanced Science* **2022**, *9*, 2200825.

(12) Zhu, X.; Ge, M.; Sun, T.; Yuan, X.; Li, Y. Rationalizing Functionalized MXenes as Effective Anchor Materials for Lithium–Sulfur Batteries via First-Principles Calculations. *The Journal of Physical Chemistry Letters* **2023**, *14*, 2215–2221.

(13) Liu, Y.; Tan, X.; Liang, J.; Han, H.; Xiang, P.; Yan, W. Machine Learning for Perovskite Solar Cells and Component Materials: Key Technologies and Prospects. *Advanced Functional Materials* **2023**, 2214271.

(14) Gorai, P.; Krasikov, D.; Grover, S.; Xiong, G.; Metzger, W. K.; Stevanović, V. A Search for New Back Contacts for CdTe Solar Cells. *Science Advances* **2023**, *9*, eade3761.

(15) Aykol, M.; Kim, S.; Hegde, V. I.; Snydacker, D.; Lu, Z.; Hao, S.; Kirklin, S.; Morgan, D.; Wolverton, C. High-Throughput Computational Design of Cathode Coatings for Li-Ion Batteries. *Nature Communications* **2016**, *7*, 13779.

(16) Sun, S.; Hartono, N. T.; Ren, Z. D.; Oviedo, F.; Buscemi, A. M.; Layurova, M.; Chen, D. X.; Ogunfunmi, T.; Thapa, J.; Ramasamy, S., et al. Accelerated Development of Perovskite-Inspired Materials via High-Throughput Synthesis and Machine-Learning Diagnosis. *Joule* **2019**, *3*, 1437–1451.

(17) Benayad, A.; Diddens, D.; Heuer, A.; Krishnamoorthy, A. N.; Maiti, M.; Cras, F. L.; Legallais, M.; Rahmanian, F.; Shin, Y.; Stein, H., et al. High-Throughput Experimentation and Computational Freeway Lanes for Accelerated Battery Electrolyte and Interface Development Research. *Advanced Energy Materials* **2022**, *12*, 2102678.

(18) Zhang, X.; Zhang, Z.; Wu, D.; Zhang, X.; Zhao, X.; Zhou, Z. Computational Screening of 2D Materials and Rational Design of Heterojunctions for Water Splitting Photocatalysts. *Small Methods* **2018**, *2*, 1700359.

(19) Lu, S.; Zhou, Q.; Guo, Y.; Zhang, Y.; Wu, Y.; Wang, J. Coupling a Crystal Graph Multilayer Descriptor to Active Learning for Rapid Discovery of 2D Ferromagnetic Semiconductors/Half-Metals/Metals. *Advanced Materials* **2020**, *32*, 2002658.

(20) Mounet, N.; Gibertini, M.; Schwaller, P.; Campi, D.; Merkys, A.; Marrazzo, A.; Sohier, T.; Castelli, I. E.; Cepellotti, A.; Pizzi, G., et al. Two-Dimensional Materials from High-Throughput Computational Exfoliation of Experimentally Known Compounds. *Nature Nanotechnology* **2018**, *13*, 246–252.

(21) Hart, G. L.; Mueller, T.; Toher, C.; Curtarolo, S. Machine Learning for Alloys. *Nature Reviews Materials* **2021**, *6*, 730–755.

(22) Feng, R.; Zhang, C.; Gao, M. C.; Pei, Z.; Zhang, F.; Chen, Y.; Ma, D.; An, K.; Poplawsky, J. D.; Ouyang, L., et al. High-Hhroughput Design of High-Performance Lightweight High-Entropy Alloys. *Nature Communications* **2021**, *12*, 4329.

(23) Li, Y.; Yang, J.; Zhao, R.; Zhang, Y.; Wang, X.; He, X.; Fu, Y.; Zhang, L. Design of Organic–Inorganic Hybrid Heterostructured Semiconductors via High-Throughput Materials Screening for Optoelectronic Applications. *Journal of the American Chemical Society* **2022**, *144*, 16656–16666.

(24) Gan, Y.; Miao, N.; Lan, P.; Zhou, J.; Elliott, S. R.; Sun, Z. Robust Design of High-Performance Optoelectronic Chalcogenide Crystals from High-Throughput Computation. *Journal of the American Chemical Society* **2022**, *144*, 5878–5886.

(25) Ling, C.; Cui, Y.; Lu, S.; Bai, X.; Wang, J. How Computations Accelerate Electrocatalyst Discovery. *Chem* **2022**, *8*, 1575–1610.

(26) Wang, S.; Gao, H.; Li, L.; San Hui, K.; Dinh, D. A.; Wu, S.; Kumar, S.; Chen, F.; Shao, Z.; Hui, K. N. High-Throughput Identification of Highly Active and Selective Single-Atom Catalysts for Electrochemical Ammonia Synthesis through Nitrate Reduction. *Nano Energy* **2022**, *100*, 107517.

(27) Li, J.; Du, P.; Li, S.; Liu, J.; Zhu, M.; Tan, Z.; Hu, M.; Luo, J.; Guo, D.; Ma, L., et al. High-Hhroughput Combinatorial Optimizations of Perovskite Light-Emitting Diodes Based on All-Vacuum Deposition. *Advanced Functional Materials* **2019**, *29*, 1903607.

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

(29) Sun, W.; Zheng, Y.; Yang, K.; Zhang, Q.; Shah, A. A.; Wu, Z.; Sun, Y.; Feng, L.; Chen, D.; Xiao, Z., et al. Machine Learning–Assisted Molecular Design and Efficiency Prediction for High-Performance Organic Photovoltaic Materials. *Science Advances* **2019**, *5*, eaay4275.

(30) Xia, R.; Brabec, C. J.; Yip, H.-L.; Cao, Y. High-Throughput Optical Screening for Efficient Semitransparent Organic Solar Cells. *Joule* **2019**, *3*, 2241–2254.

(31) Westermayr, J.; Gilkes, J.; Barrett, R.; Maurer, R. J. High-Throughput Property-Driven Generative Design of Functional Organic Molecules. *Nature Computational Science* **2023**, 1–10.

(32) Behrendt, D.; Banerjee, S.; Clark, C.; Rappe, A. M. High-Throughput Computational Screening of Bioinspired Dual-Atom Alloys for CO_2 Activation. *Journal of the American Chemical Society* **2023**, *145*, 4730–4735.

(33) Tang, F.; Ono, S.; Wan, X.; Watanabe, H. High-Throughput Investigations of Popological and Nodal Superconductors. *Physical Review Letters* **2022**, *129*, 027001.

(34) Chen, Y.; Bai, X.; Liu, D.; Fu, X.; Yang, Q. High-Throughput Computational Exploration of MOFs with Open Cu Sites for Adsorptive Separation of Hydrogen Isotopes. *ACS Applied Materials & Interfaces* **2022**, *14*, 24980–24991.

(35) Lischka, H.; Nachtigallova, D.; Aquino, A. J.; Szalay, P. G.; Plasser, F.; Machado, F. B.; Barbatti, M. Multireference Approaches for Excited States of Molecules. *Chemical Reviews* **2018**, *118*, 7293–7361.

(36) Axelrod, S.; Shakhnovich, E.; Gómez-Bombarelli, R. Excited State Non-Adiabatic Dynamics of Large Photoswitchable Molecules Using a Chemically Transferable Machine Learning Potential. *Nature Communications* **2022**, *13*, 3440.

(37) Westermayr, J.; Marquetand, P. Machine Learning for Electronically Excited States of Molecules. *Chemical Reviews* **2020**, *121*, 9873–9926.

(38) Rong, Y.; Bian, Y.; Xu, T.; Xie, W.; Wei, Y.; Huang, W.; Huang, J. Self-Supervised Graph Transformer on Large-Scale Molecular Data. *Advances in Neural Information Processing Systems* **2020**, *33*, 12559–12571.

(39) Wang, Y.; Wang, J.; Cao, Z.; Barati Farimani, A. Molecular Contrastive Learning of Representations via Graph Neural Networks. *Nature Machine Intelligence* **2022**, *4*, 279–287.

(40) Kenton, J. D. M.-W. C.; Toutanova, L. K. Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding. Proceedings of NAACL-HLT. 2019; pp 4171–4186.

(41) Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; Sutskever, I., et al. Language Models Are Unsupervised Multitask Learners. *OpenAI Blog* **2019**, *1*, 9.

(42) Ying, C.; Cai, T.; Luo, S.; Zheng, S.; Ke, G.; He, D.; Shen, Y.; Liu, T.-Y. Do Transformers Really Perform Badly for Graph Representation? *Advances in Neural Information Processing Systems* **2021**, *34*, 28877–28888.

(43) Zhou, G.; Gao, Z.; Ding, Q.; Zheng, H.; Xu, H.; Wei, Z.; Zhang, L.; Ke, G. Uni-Mol: A Universal 3D Molecular Representation Learning Framework. The Eleventh International Conference on Learning Representations. 2023.

(44) Li, T.-Y.; Wu, J.; Wu, Z.-G.; Zheng, Y.-X.; Zuo, J.-L.; Pan, Y. Rational Design of

Phosphorescent Iridium (III) Complexes for Emission Color Tunability and Their Applications in OLEDs. *Coordination Chemistry Reviews* **2018**, *374*, 55–92.

(45) Turcani, L.; Tarzia, A.; Szczypiński, F. T.; Jelfs, K. E. stk: An Extendable Python Framework for Automated Molecular and Supramolecular Structure Assembly and Discovery. *The Journal of Chemical Physics* **2021**, *154*, 214102.

(46) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *Journal of Chemical Theory and Computation* **2019**, *15*, 1652–1671.

(47) Frisch, M. J. et al. Gaussian˜16 Revision C.01. 2016; Gaussian Inc. Wallingford CT.

(48) Neese, F. The ORCA Program System. *WIREs Computational Molecular Science* **2012**, *2*, 73–78.

(49) Neese, F. Software Update: the ORCA Program System, Version 4.0. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2018**, *8*, e1327.

(50) Shuai, Z.; Peng, Q. Organic Light-Emitting Diodes: Theoretical Understanding of Highly Efficient Materials and Development of Computational Methodology. *National Science Review* **2017**, *4*, 224–239.

(51) Peng, Q.; Yi, Y.; Shuai, Z.; Shao, J. Toward Quantitative Prediction of Molecular Fluorescence Quantum Efficiency: Role of Duschinsky Rotation. *Journal of the American Chemical Society* **2007**, *129*, 9333–9339.

(52) Shuai, Z. Thermal Vibration Correlation Function Formalism for Molecular Excited State Decay Rates. *Chinese Journal of Chemistry* **2020**, *38*, 1223–1232.

(53) Chao, K.; Shao, K.; Peng, T.; Zhu, D.; Wang, Y.; Liu, Y.; Su, Z.; Bryce, M. R. New Oxazoline-and Thiazoline-Containing Heteroleptic Iridium (III) Complexes for Highly-

Efficient Phosphorescent Organic Light-Emitting Devices (PhOLEDs): Colour Tuning by Varying the Electroluminescence Bandwidth. *Journal of Materials Chemistry C* **2013**, *1*, 6800–6806.

(54) Wang, R.; Liu, D.; Ren, H.; Zhang, T.; Wang, X.; Li, J. Homoleptic Tris-Cyclometalated Iridium Complexes with 2-Phenylbenzothiazole Ligands for Highly Efficient Orange OLEDs. *Journal of Materials Chemistry* **2011**, *21*, 15494–15500.

(55) Ma, X.; Liang, J.; Bai, F.; Ye, K.; Xu, J.; Zhu, D.; Bryce, M. R. New Mixed-C^N Ligand Tris-Cyclometalated IrIII Complexes for Highly-Efficient Green Organic Light-Emitting Diodes with Low Efficiency Roll-Off. *European Journal of Inorganic Chemistry* **2018**, *42*, 4614–4621.

(56) Lee, J.; Park, H.; Park, K.-M.; Kim, J.; Lee, J.-Y.; Kang, Y. Deep-Blue Phosphorescent Iridium (III) Dyes Based on Fluorine-Functionalized bis (2,3-Bipyridyl) Ligand for Efficient Organic Light-Emitting Diodes. *Dyes and Pigments* **2015**, *123*, 235–241.

(57) Zhao, J.; Yu, Y.; Yang, X.; Yan, X.; Zhang, H.; Xu, X.; Zhou, G.; Wu, Z.; Ren, Y.; Wong, W.-Y. Phosphorescent Iridium (III) Complexes Bearing Fluorinated Aromatic Sulfonyl Group with Nearly Unity Phosphorescent Quantum Yields and Outstanding Electroluminescent Properties. *ACS Applied Materials & Interfaces* **2015**, *7*, 24703–24714.

(58) Jing, Y.-M.; Zheng, Y. Photoluminescence and Electroluminescence of an Iridium(III) Complex with 2,6-Bis(Trifluoromethyl)-2,4-Bipyridine and 2-(5-Phenyl-1,3,4-Thiadiazol-2-yl)Phenol Ligands. *New Journal of Chemistry* **2017**, *41*, 3029–3035.

(59) Jou, J.-H.; Lin, Y.-X.; Peng, S.-H.; Li, C.-J.; Yang, Y.-M.; Chin, C.-L.; Shyue, J.-J.; Sun, S.-S.; Lee, M.; Chen, C.-T., et al. Highly Efficient Yellow Organic Light Emitting Diode with a Novel Wet-and Dry-Process Feasible Iridium Complex Emitter. *Advanced Functional Materials* **2014**, *24*, 555–562.

(60) Tan, G.; Chen, S.; Sun, N.; Li, Y.; Fortin, D.; Wong, W.-Y.; Kwok, H.-S.; Ma, D.; Wu, H.; Wang, L., et al. Highly Efficient Iridium (III) Phosphors with Phenoxy-Substituted Ligands and Their High-Performance OLEDs. *Journal of Materials Chemistry C* **2013**, *1*, 808–821.

(61) Li, T.-Y.; Liang, X.; Zhou, L.; Wu, C.; Zhang, S.; Liu, X.; Lu, G.-Z.; Xue, L.-S.; Zheng, Y.-X.; Zuo, J.-L. N-heterocyclic Carbenes: Versatile Second Cyclometalated Ligands for Neutral Iridium (III) Heteroleptic Complexes. *Inorganic Chemistry* **2015**, *54*, 161–173.

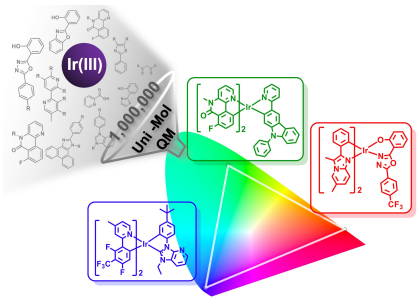
(62) Joung, J. F.; Han, M.; Jeong, M.; Park, S. Experimental Database of Optical Properties of Organic Compounds. *Scientific Data* **2020**, *7*, 295.

(63) Joung, J. F.; Han, M.; Hwang, J.; Jeong, M.; Choi, D. H.; Park, S. Deep Learning Optical Spectroscopy Based on Experimental Database: Potential Applications to Molecular Design. *JACS Au* **2021**, *1*, 427–438.

(64) Adachi, C.; Kwong, R. C.; Djurovich, P.; Adamovich, V.; Baldo, M. A.; Thompson, M. E.; Forrest, S. R. Endothermic Energy Transfer: A Mechanism for Generating Very Efficient High-Energy Phosphorescent Emission in Organic Materials. *Applied Physics Letters* **2001**, *79*, 2082–2084.

(65) Ryu, B.; Kim, K.; Ha, Y.; Bae, J.; Lee, S.; Song, J.; Lee, K.; Lee, J.; Kim, K.; Kim, H. New RGB Primary for Various Multimedia Systems. *Journal of Information Display* **2014**, *15*, 65–70.

(66) Smet, K.; Schanda, J.; Whitehead, L.; Luo, R. CRI2012: A Proposal for Updating the CIE Colour Rendering Index. *Lighting Research & Technology* **2013**, *45*, 689–709.



TOC Graphic