A general protocol for phosphorescent platinum(II) complexes: generation, high throughput virtual screening and highly accurate predictions

OPEN ACCESS
IOP Publishing | Songshan Lake Materials Laboratory

MATERIALS FUTURES

Mater. Futures 4 (2025) 025601 (10pp)

https://doi.org/10.1088/2752-5724/adb320

A general protocol for phosphorescent platinum(II) complexes: generation, high throughput virtual screening and highly accurate predictions

Shuai Wang[1], ChiYung Yam[2,3,*], LiHong Hu[4], Faan-Fung Hung[1,2,5], Shuguang Chen[1,2], Chi-Ming Che[1,2,5,*] and GuanHua Chen[1,2,*]

[1] Department of Chemistry, The University of Hong Kong, Pokfulam, Hong Kong Special Administrative Region of China, People's Republic of China
[2] Hong Kong Quantum AI Lab Limited, Pak Shek Kok, Hong Kong Special Administrative Region of China, People's Republic of China
[3] Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China, Shenzhen 518000, People's Republic of China
[4] School of Information Science and Technology, Northeast Normal University, Changchun 130117, People's Republic of China
[5] State Key Laboratory of Synthetic Chemistry, HKU-CAS Joint Laboratory on New Materials, The University of Hong Kong, Pokfulam, Hong Kong Special Administrative Region of China, People's Republic of China

E-mail: ghc@everest.hku.hk, yamcy@uestc.edu.cn and cmche@hku.hk

Received 30 September 2024, revised 26 December 2024
Accepted for publication 16 January 2025
Published 4 April 2025

Abstract

The utilization of phosphorescent metal complexes as emissive dopants for organic light-emitting diodes (OLEDs) has been the subject of intense research. Cyclometalated Pt(II) complexes are particularly popular triplet emitters due to their color-tunable emissions. To make them viable for practical applications as OLED emitters, it is essential to develop Pt(II) complexes with high radiative decay rate constants (k_r) and photoluminescence quantum yields (PLQY). To this end, an efficient and accurate prediction tool is highly desirable. In this work, we propose a general yet powerful protocol achieving metal complex generation, high throughput virtual screening (HTVS), and fast predictions with high accuracy. More than 3600 potential structures are generated in a synthesis-friendly manner. Moreover, three HTVS-machine learning (ML) models are established using different algorithms with carefully designed features that are suitable for metal complexes. Specifically, 30 potential candidates are filtered out by HTVS-ML models with a three-tier screening rule and put into accurate predictions with experimental calibration Δ-learning method. The highly accurate prediction approach further reduces the stress of experiments and inspires greater confidence in identifying

* Authors to whom any correspondence should be addressed.

the most promising complexes as excellent emitters. As a result, 12 promising complexes ($k_r > 10^5$ s^{-1} and PLQY > 0.6) with the same superior core structures are confirmed from over 3600 Pt-complexes. Experiments revealed that two very close complexes have excellent emission properties and are consistent with the prediction results, providing strong evidence for the efficacy of the proposed protocol. We expect this protocol will become a valuable tool, expediting the rational design and rapid development of novel OLED materials with desired properties.

Supplementary material for this article is available online

Keywords: phosphorescent emitters, OLED, generation, high throughput virtual screening, machine learning

1. Introduction

Organic light-emitting diodes (OLEDs) are becoming increasingly popular as sustainable light sources in digital displays, portable systems, and many other fields [1]. While the first-generation fluorescence-based OLEDs are limited to 25% internal quantum efficiency (IQE) due to the 1:3 ratio of singlet and triplet excitons according to spin statistics [2]. To overcome this limitation, triplet excitons can be utilized through the use of phosphorescent heavy metal-based emitters such as iridium and platinum. These heavy metal atoms can induce strong spin–orbit coupling to facilitate the intersystem crossing process from the singlet to triplet excited states and to promote radiative decay from the triplet excited state through phosphorescence [3]. As a result, the second-generation phosphorescent OLEDs (PhOLEDs) based on iridium/platinum emitters can achieve IQE up to 100%. Over the past decade, research on platinum-based PhOLEDs has grown significantly and device performances have been improved with the introduction of tetradentate cyclometalated platinum complexes [4–8]. However, it would still require substantial time and cost to develop superior PhOLED emitters in a trial-and-error manner through experiments.

Density functional theory [9] (DFT) and time-dependent DFT [10] (TDDFT) are extensively employed to predict material properties [11–13]. These methods provide a good balance between accuracy and efficiency when simulating the properties of target systems. Nevertheless, despite the advantages of first-principles calculations, there are certain limitations and challenges [14–16]. Notably, the results derived from DFT or TDDFT calculations on phosphorescent platinum emitters are still not accurate compared with the experimental ones [17]. Besides, the computational costs are usually not affordable for a large library of metal complexes when exploring vast chemical space.

High throughput virtual screening (HTVS) [18, 19] with machine learning (ML) algorithms has become a vital technology in pharmaceutical research [20–24] and material development [25–29], which rapidly evaluates potential candidates from a large library of compounds. Despite the considerable investments made into HTVS-ML, there remain several daunting challenges in the Pt-based PhOLED field. These include the lack of experimental/computational data for platinum PhOLED emitters, as well as the absence of potential structures in public datasets. Furthermore, the utilization of simplified molecular-input line-entry system (SMILES) [30] is not suitable due to the presence of transition metals and coordination bonds in metal complexes. Additionally, popular descriptors/fingerprints are not effective in this context due to the complicated situations caused by multiple coordination bonds in molecules. In terms of data, many public datasets [31–35] have been developed, but few of them contain transition metals regardless of the experimental/computational results. David Balcells [36] proposed a dataset with different transition metals in simple coordination ways, though the Pt-complexes therein cannot be considered as potential phosphorescent emitters. Additionally, many generation algorithms based on SMILES and graphs are not easily implemented in the Pt-emitters field. Nevertheless, Maestro in Schrodinger package [37] offers a means of expanding the chemical space for potential emitter candidates, particularly those involving transition metals. While PDB and MOL formats are suitable for representing Pt-complexes, they still require meticulous modification and verification to generate accurate descriptors and fingerprints. Generally, the popular descriptors/fingerprints extracted from the RDKit [38] Python library are effective for pure organic molecules, regardless of the size of molecules, the types of contained atoms, and the different coordination bonds. However, for Pt-complexes, using the popular descriptors without proper revision and check can lead to significant errors, even in the simplest descriptors such as molecular weights. Therefore, modified and self-designed descriptors are employed to represent the complexes and their external environments.

After addressing the aforementioned issues, HTVS-ML models are established to initially predict phosphorescence-related properties such as the energy gaps between the lowest triplet state (T_1) and the ground state (S_0), as well as between T_1 and the first singlet state (S_1), along with the non-radiative decay rate constant (k_{nr}). Notably, the capability and generality of the models are evaluated by not only the independent testing set but an external set consisting of recently reported Pt-complexes. Furthermore, a three-tier screening strategy is devised to filter out promising candidates from both radiation

and non-radiation perspectives. As a result, potential complexes are rapidly identified by virtual screening and they are highly expected to exhibit superior performance as emitters. While an accurate assessment of emission wavelength, radiative decay rate constant (k_r), and photoluminescence quantum yield (PLQY) is paramount, an approach that combines first-principles quantum mechanical method, ML and experimental calibration is introduced and implemented to obtain highly precise prediction results [39]. This approach further reduces the burden of synthesis by confirming the photophysical properties of the emitters. The workflow of the proposed protocol is illustrated in figure 1(a). The promising ones are eventually confirmed with $k_r > 10^5$ s^{-1} and PLQY > 0.6 via highly accurate predictions. It is noteworthy that several of these promising candidates exhibit striking similarities to the latest synthesized samples prepared by the experimental group, attesting to the remarkable efficiency and accuracy of the proposed protocol. In summary, this protocol, which incorporates molecule generation, HTVS and highly accurate predictions, provides an efficient tool for the design and advancement of novel OLED materials.

2. Methodologies

2.1. Dataset construction and division

In this work, our primary focus is on cyclometalated Pt(II) complexes with tridentate or tetradentate ligands. Photophysical data of 198 phosphorescent Pt-complexes reported in the literature are collected, including emission wavelength/energy, lifetime, PLQY, radiative decay rate constant (k_r) and non-radiative decay rate constant (k_{nr}). These were mostly measured in ambient conditions and degassed solutions [5], the distributions of which are shown in figures 1(b)–(d). Besides, S_1-T_1 gap is calculated by PBE0 under TZP basis set using ADF package [40]. The details of data pre-processing can be found in section 2 of Supporting Information (SI). The emission energy spans a range from 1.88 to 2.88 eV, with a mean value of 2.39 eV. The S_1-T_1 gap of the compounds distributes with a mean value of 0.456 eV except for several extremely high values. On the other hand, k_{nr} values span over seven orders of magnitude, with a majority spanned between 10^5 s^{-1} and 10^7 s^{-1} and a mean value of 1.03×10^6 s^{-1}. This poses a great difficulty in constructing ML models for k_{nr} prediction. To address this issue, meticulous care has been taken in the data processing with log and data division. An improved Kennard Stone algorithm [41], which partitions the sample set based on maximum–minimum X–Y distance [42], is employed to ensure a relative balance between training and testing sets within the limited and unbalanced data. The distance can be calculated using equation (1).

$$d_{xy}(p,q) = \frac{d_x(p,q)}{\max_{p,q\in[1,N]} d_x(p,q)} + \frac{d_y(p,q)}{\max_{p,q\in[1,N]} d_y(p,q)} \quad (p,q \in [1,N]) \tag{1}$$

where x and y represent the features and the target properties, respectively. p and q denote pth and qth samples in the whole dataset with N samples.

2.2. Features for HTVS-ML

Although PDB and MOL formats may help to get a molecule read by RDKit [38] library, there are always issues such as the count of aromatic rings and even the molecular weight. Therefore, it is essential to carry out careful bond modification, selection of calculated descriptors, and creation of new proper descriptors in order to obtain correct and valid descriptors/fingerprints for the Pt- complexes. Basically, we employ a strategy considering core structures, molecular properties and testing medium to represent each complex. In the core structure part, due to the great differences and numerous variations, a set of descriptors has been designed to quickly distinguish and identify the core part. The descriptors that depend on the empirical pairwise Pauling electronegativity difference ($\Delta\chi$) between the coordination atom and any ith adjacent atom apart from Pt atom are presented:

$$\Delta\chi_{\text{coor},i} = \chi_{\text{coor}} - \chi_i. \tag{2}$$

These core structural differences include the maximum ($\Delta\chi_{\max}$), and minimum ($\Delta\chi_{\min}$), as well as the sum:

$$\Delta\chi_{\text{sum}} = \sum_{j\in\text{lig.}} \sum_{i\in\text{AA}} \Delta\chi^j_{\text{coor},i} \tag{3}$$

which is taken over the coordination atom and all atoms adjacent (AA) to it for all ligands (lig.) in the molecule. Regarding the molecular properties part, 85 descriptors are cautiously selected from MoleculeDescriptors module from RDKit [38] library (version 2020.09). These descriptors predominantly comprise molecular information, including the count of aromatic nitrogen and rotatable bonds, the proportion of C atoms that are sp^3 hybridized, and the average molecular weight of the molecule without hydrogens. In addition, ring information such as the number of benzene rings, furan rings, aliphatic heterocycles, aromatic carbocycles, and saturated carbocycles, as well as the Kappa [43] and Chi [44] series topological descriptors, are also included. Hence a comprehensive description of the complexes is accomplished. Additionally, the environments in which the Pt-complex is tested have a significant impact on the photophysical properties. Consequently, general environment descriptors were developed to better present the solvent effect and enable rapid ML predictions. These descriptors relate to the testing medium and further details can be found in section 3 of SI.

2.3. Feature engineering

In total, 153 descriptors are created for Pt-complexes. To avoid overfitting and low generality, feature selection and importance analysis are implemented while training the models. All the features are firstly filtered by variance with a threshold of 0, which means no difference for the training data. Then,

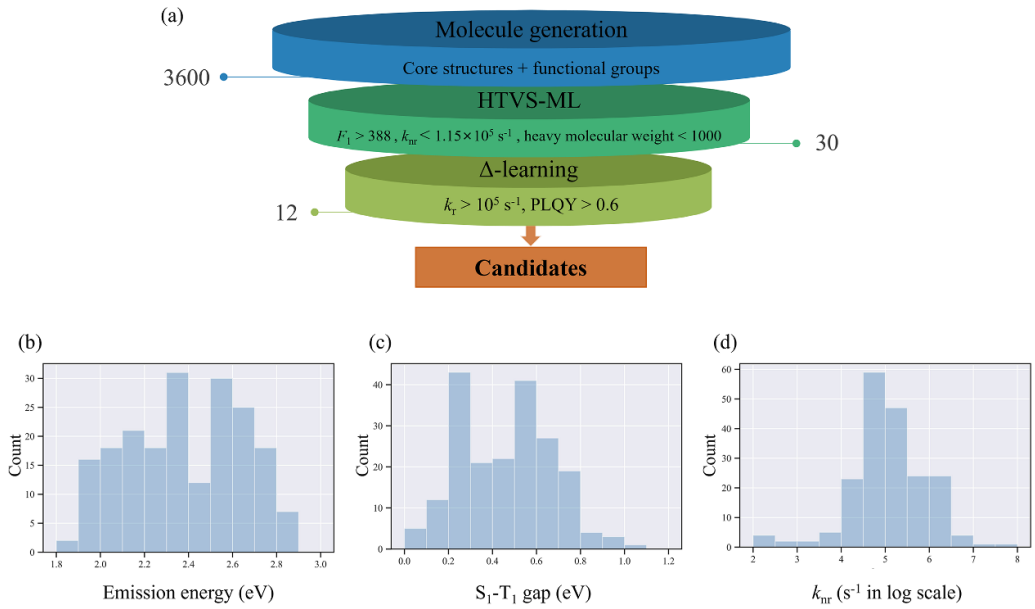

Figure 1. (a) Workflow of the proposed protocol; Distribution of (b) emission energy, (c) S_1-T_1 gap and (d) non-radiative decay rate constant (k_{nr}) in the dataset.

recursive feature elimination (RFE) based on random forest [45] algorithm is employed to rapidly reduce the complexity of the model by recursively removing features that are not contributing to the targets. RFE assists in identifying the important features that are most relevant to the target variable. Subsequently, feature importance analysis after RFE helps to confirm the most important features, and all the features with no contributions are removed from the feature matrix. The feature selection details are illustrated in Section 4 of SI. As a result, the optimal HTVS-ML models are trained and benefit the screening procedure.

2.4. ML algorithms

Six types of algorithms including adaptive boosting [46], light gradient boosting (GB) machine [47] (LightGBM), extreme GB[48] (XGB), support vector machines [49] (SVM), k-nearest neighbors [50] (KNN), and kernel ridge regression [51] (KRR) are considered to establish regression models. After data division, 80% of the dataset was partitioned for model training, and the remaining 20% was taken as an independent test set. The hyperparameters are tuned by Python library hyperopt [52] with 10-fold cross-validation. Details of the hyperparameter optimization can be found in Section 4 of SI. Performances of the models are evaluated based on the correlation coefficient (r), the root mean square error (RMSE), and mean absolute error (MAE). Partial features employed in HTVS-ML models are calculated by RDKit [38] package in Python environment and the designed descriptors can be obtained through Data and code availability section.

2.5. Screening rules

Based on the HTVS-ML models, three crucial properties are predicted and can be set as screening criteria. Generally, radiative decay rate constant (k_r) and PLQY (ϕ_{PL}) are both key factors when developing new phosphorescent emitters and PLQY can be calculated as follows:

$$\phi_{PL} = \frac{k_r}{k_r + k_{nr}} = \frac{1}{1 + \frac{k_{nr}}{k_r}}. \tag{4}$$

It is found that k_r and k_{nr} are with enormous significance for Pt-complexes, thus screening excellent emitters can be conducted by increasing k_r and decreasing k_{nr}. In other words, higher k_r and lower k_{nr} contribute to the outstanding performance of the phosphorescent emitters. A three-tier screening rule is rationally designed to filter out the promising candidates from both radiation and non-radiation aspects.

Tier 1: Radiative decay rate constant k_r is a fundamental quantity that determines the performance of OLEDs. Low k_r results in high efficiency roll-off at high luminance and material degradation due to side reactions. The radiative decay rate constant (k_r) can be calculated as follows:

$$k_r(\tilde{\nu}) = \frac{8\pi^2\eta^3\tilde{\nu}^3}{3\varepsilon_0\hbar}|M|^2 \tag{5}$$

where ν is the emission energy from the lowest triplet state (T_1) to the ground state (S_0); ε_0 and $\hbar$ denote the vacuum permittivity and reduced Planck constant, respectively; η represents the refractive index and M stands for the transition dipole moment of $T_1 \rightarrow S_0$ transition. In the framework of perturbation theory and assuming a three-level system, the transition dipole moment M that couples the emitting triplet state to the singlet ground state is approximated by:

$$M \approx \frac{\langle \psi_{T_1} | H_{SOC} | \psi_{S_1} \rangle}{E_{S_1} - E_{T_1}} M_{S_1} \tag{6}$$

where M_{S_1} is the transition dipole moment of the perturbing singlet state. According to the expressions (5) and (6), the

score function F_1 can be set as below:

$$F_1 = \frac{\nu^3}{(E_{S_1} - E_{T_1})^2} \tag{7}$$

where the emission energy and S_1-T_1 energy gap can be promptly predicted using the HTVS-ML models. F_1 value is proportional to k_r thus implying the possibility of high k_r (neglecting other parameters such as H_{SOC} and M_{s_1} at this stage).

Tier 2: Having a high possibility of superior k_r is not enough to guarantee the exceptional performance of the emitters. k_{nr} is taken into consideration and evaluated in this tier. After selecting the complexes with higher k_r possibility in Tier 1, low k_{nr} complexes are sought by the HTVS-ML model.

Tier 3: Excel out the complexes with heavy molecular weights of more than 1000, which may not be friendly to the sublimation process for purification and device fabrication via vapor deposition. Through the three-tier screening, promising candidates can be filtered out and put into a high-accuracy prediction step. The specific criteria of the three tiers and the rationale behind them are illustrated in section 8 of SI.

2.6. Accurate predictions via Δ-learning

A combination of first-principles quantum mechanical method, ML and experimental calibration is employed to obtain highly accurate prediction results. Our previous results have demonstrated that ensemble learning models combined with stacking-based approaches yield the best performance [39]. First-principles simulations are utilized to calculate the photophysical properties of complexes to generate molecular features. Overall, we select the features that are associated with the metal coordination, as well as those related to the photophysical properties of the Pt-complexes, for training the ML models. In a Pt-emitter, the metal ion and its surrounding atoms play key roles in the phosphorescence process, thus the average electron densities of these atoms are taken as features. Additionally, the coordination bond type and coordination bond length are considered as well. Furthermore, the calculated emission energy, oscillator strength, and spin–orbit coupling constants are included in the feature matrix. Details on first-principles calculations can be found in section 6 of SI or Methodologies section of [39].

3. Results and discussions

3.1. Potential complex generation

From a synthesis-friendly perspective, four core structures and fifteen functional groups are selected from the literature. The H atoms at substitution sites (highlighted in figure 2) are replaced with various functional groups. For symmetric core structures, the substitution sites are kept in a symmetric manner (core I, II, and III) with up to 4 substitutions. For the core IV, the substitutions are limited to 2 sites and cannot occur in the same 6-membered ring. As a result, more than 3600 complexes are generated using the Schrödinger package [37] with PDB format which can be quickly read by RDKit [38] python library.

3.2. HTVS

3.2.1. Rapid screening ML models. Approximately 200 phosphorescent platinum(II) complexes have been collected and employed as the dataset for ML. In this study, we develop three HTVS-ML models to predict emission energy, S_1-T_1 gap, and non-radiative decay rate constant (k_{nr}). The ML models are constructed using various algorithms, including AdaBoost [46], LightGBM [47], XGB [48], SVM [49], KNN [50] and KRR [51]. The performance of each model in predicting emission energy is detailed in table S4, while the optimal one is summarized in table 1. LightGBM demonstrates the best performance in terms of both correlation coefficient and error metrics. This algorithm enhances the efficiency and scalability of GB algorithm without compromising its inherent effectiveness, making it suitable for the rapid evaluation of a large library of potential molecules. Subsequently, feature selection is conducted using variance filter, RFE and feature importance analysis. The results are presented in table S7. Ultimately, a total of 48 features are selected, with the predicted results and the ten most significant features illustrated in figures 3(a) and 4(a), respectively. FpDensityMorgan [53] series descriptors contribute significantly to the model, ranking first, third, and sixth. FpDensityMorgan generates the similarity fingerprints using certain chemical and connectivity attributes of atoms, which are essentially a normalized count of the number of unique atomic environments in a molecule, which can be increased/decreased by adding/removing unique substructures. Besides, HeavyAtomMolWt and BalabanJ [54], which calculate the molecular weight of the molecule ignoring hydrogens, and Balaban's J value for a molecule are in the second and fourth place among all the features. Additionally, topological descriptors like Kappa1 [43] and Kappa3 [43] also contribute to the optimal model.

On the side of S_1-T_1 gap, the optimal model's performance following 10-fold cross-validation is listed in table 1. Similarly, LightGBM achieves the best performance. Furthermore, feature engineering and importance analysis, detailed in table S8, highlight the most important features depicted in figure 4(b). The BalabanJ [54] value presents the most remarkable contribution in determining the target property. In addition, structure-related descriptors such as the numbers of bicyclic rings, *para*-hydroxylation sites, and aliphatic heterocycles contribute significantly to the model as well. Others are primarily similarity fingerprints and topological descriptors. The prediction results of S_1-T_1 gap based on the optimal model can be seen in figure 3(b).

Non-radiative decay rate constant k_{nr} is a fundamental property that determines the efficiency of OLEDs. Emitters with large k_{nr} values suffer from low PLQYs, which would

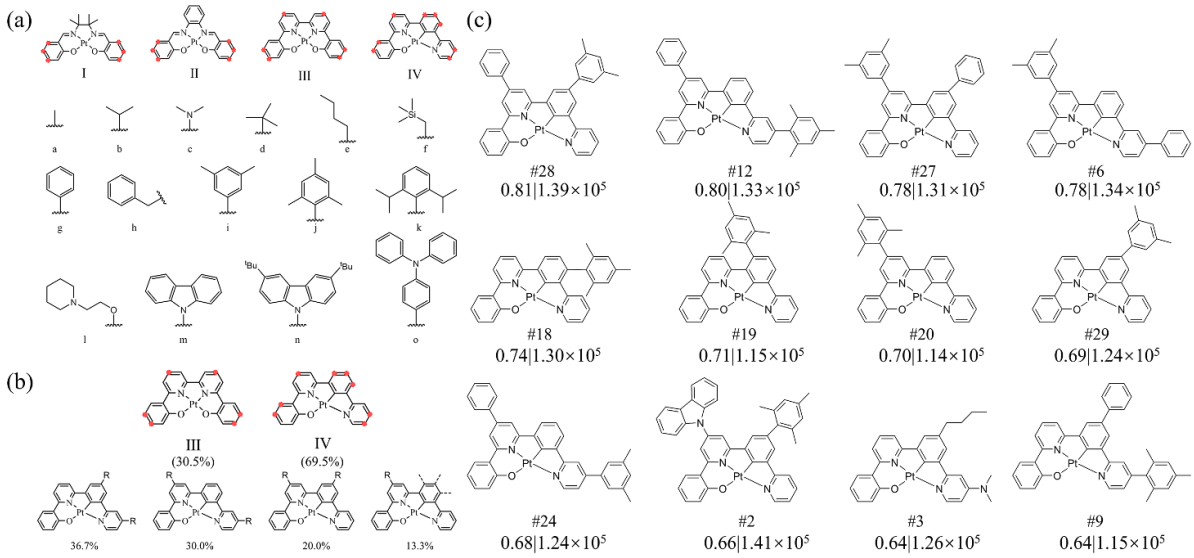

Figure 2. (a) Four core structures and fifteen functional groups for Pt-complexes generation (b) HTVS statistical collection of 200 selected complexes after Tier 1 (top) and 30 potential candidates after Tier 2 & 3 (bottom) (c) Promising candidates with accurate PLQY (left) and k_r (right) prediction results.

Table 1. Performance of the optimal ML models in predicting emission energy, S_1-T_1 gap and k_{nr}.

		Independent testing set[a]		
Tasks	Optimal ML models	MAE[b]	RMSE[b]	r
Emission energy	LightGBM	0.06 ± 0.01	0.09 ± 0.01	0.95 ± 0.01
S_1-T_1 gap	LightGBM	0.07 ± 0.01	0.11 ± 0.01	0.77 ± 0.04
k_{nr}	SVM	0.38 ± 0.01	0.47 ± 0.02	0.67 ± 0.04

[a] The standard deviations are calculated by the difference in the prediction of each fold.
[b] The errors for energy terms are measured in eV, while k_{nr} is measured in s^{-1} in log scale.

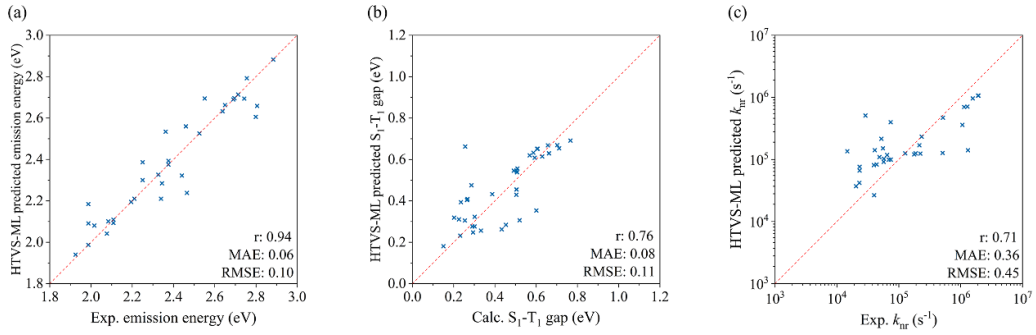

Figure 3. HTVS-ML model performances of (a) emission energy, (b) S_1-T_1 gap, and (c) non-radiative decay rate constant (k_{nr}) on the independent testing set.

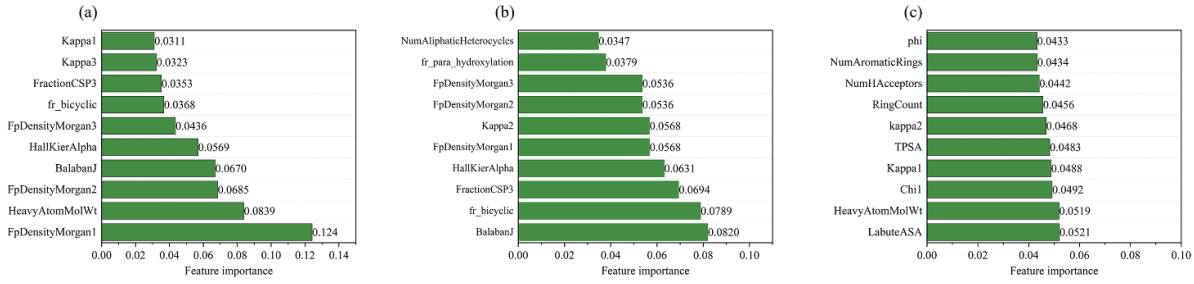

Figure 4. Ten most important features for (a) emission energy extracted from LightGBM-based ML model (b) S_1-T_1 gap extracted from LightGBM-based ML model (c) k_{nr} extracted from SVM-based ML model.

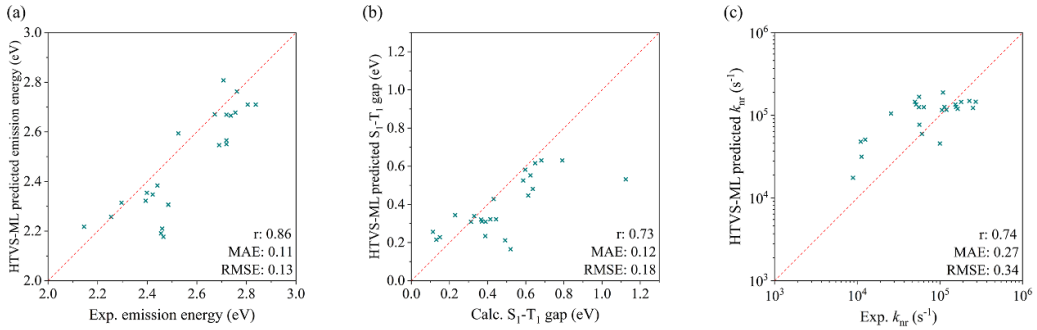

Figure 5. HTVS-ML model performances of (a) emission energy, (b) S_1-T_1 gap, and (c) non-radiative decay rate constant (k_{nr}) on the external testing set.

also exhibit low electric-to-light conversion efficiencies in OLEDs. To efficiently predict the k_{nr} values, six models are constructed using various ML algorithms, with their performance summarized in table S6. It is observed that SVM and LightGBM exhibit comparable correlation coefficients and MAEs. However, SVM demonstrates the lowest RMSE and is therefore selected as the optimal model. Feature engineering and analysis are performed, resulting in the ten most important features, which are illustrated in table S9 and figure 4(c), respectively. LabuteASA [55] and TPSA [56], which represent Labute's approximate surface area and topological polar surface area based on fragments, respectively, have superior importance compared to other features. Molecular weight without hydrogens is the second most important descriptor, followed by the number of rings, H acceptors, and aromatic rings. Additionally, topological descriptors such as Chi [44] and Kappa [43] are also important features in the model. The performance of the optimal model for k_{nr} is plotted in figure 3(c).

Overall, the emission energy model demonstrates exceptional performance in accurately evaluating target molecules, with MAE and RMSE capped at 0.1 eV. Notably, the r values for both S_1-T_1 gap and k_{nr} exceed 0.7, indicating strong precision and suitability for high-throughput virtual screening. Additionally, the generality of the three models is further assessed using an external dataset comprising recently reported Pt-complexes. The performance is presented in figure 5. For emission energy, the r value for the external set reaches as high as 0.86. Furthermore, the errors observed in the external dataset for the S_1-T_1 gap and k_{nr} are comparable to, or even superior to, those found in the independent testing set. This highlights the model's superior capability in screening the generated complexes, with further details provided in section 5 of the SI.

The utilization of the developed HTVS-ML models for emission energy and the S_1-T_1 gap facilitates the efficient screening of complexes with high F_1 scores, which are likely to exhibit elevated k_r values, based on the screening criteria outlined in Tier 1. Furthermore, the emission energy can be converted into emission wavelength, aiding in the preliminary differentiation of the emission color. Following the Tier 1 screening, the HTVS-ML model for k_{nr} can be applied to identify complexes with high PLQY, as indicated in Tier 2. By incorporating considerations of synthesis, we can effectively select promising candidates with the support of the three HTVS-ML models.

3.2.2. HTVS results. With the three optimal HTVS-ML models and well-designed screening rules, 200 molecules with the largest F_1 score are filtered out according to the screening rule in Tier 1. Among these complexes, 69.5% are substituted on the core IV as presented in figure 2(b). In comparison, 30.5% are based on core III which initially presents the superiority of core structure IV in terms of the high probability of excellent k_r performance. Moreover, the highest-performing complex based on core III ranks only 67th among 200 molecules. Following Tier 2 and 3 screening, 30 candidate complexes with the lowest k_{nr} are selected, indicating a potentially higher PLQY. See complex structures and prediction results from HTVS models in figure S2 and table S12 of SI. It can be observed that they are all on the basis of core IV and 86.7% of them are with 2-site substitutions. The promising ones are subsequently put into a highly accurate prediction step.

3.3. Accurate predictions

Although 30 promising complexes with a high potential for superior emitter performance have been screened, there is a lack of a direct and highly accurate assessment of their emission properties. In previous work [39], a general protocol was developed for accurate predictions of emission wavelength, radiative decay rate constant (k_r), and PLQY for Pt(II) phosphorescent emitters, based on the combination of first-principles quantum mechanical method, ML, and experimental calibration. Ensemble learning models combined with stacking-based approaches exhibited the best performance, and their generality for a broad palette of Pt(II) emitters has been proven. Employing the three accurate ML models, emission wavelength, k_r, and PLQY of the 30 complexes are predicted and listed in table S13 of SI. Among them, 86.7% of the complexes own a PLQY over 0.5, with a mean value of 0.60, and 76.7% of the molecules are with more than 9×10^4 s^{-1} in terms of k_r with a mean value of 1.02×10^5 s^{-1}, which

strongly demonstrates the efficiency and validity of the HTVS-ML models and screening rules. As a result, 12 complexes are eventually selected with the criteria of PLQY > 0.6 and $k_r > 10^5$ s^{-1} and are recommended for further synthesis and testing. The structures and prediction results can be seen in figure 2(c) in a PLQY descending order. The indexes in figure 2(c) are the ranks after HTVS. Among these 12 complexes, the highest PLQY and k_r are 0.81 and 1.41×10^5 s^{-1}, respectively. The molecule 20 is very close to the complex [5] which has been developed previously with a PLQY of 0.73 and a k_r of 1.55×10^5 s^{-1}. Moreover, molecule 12 closely resembles the structure of a new complex (Pt-1 in the [57]) with PLQY over 0.80 and k_r 3×10^5 s^{-1}. Minor discrepancies between the predicted results and experimental ones can be observed, highlighting the excellent capabilities of the HTVS-ML models and highly accurate prediction models. The dataset expansion and more experimental comparisons are presented in section 9 of SI. In summary, the highly accurate prediction approach not only reduces the experimental stress, but also provides greater confidence in identifying the most promising complexes as exceptional emitters.

4. Conclusion

In summary, a general protocol of platinum complex generation, HTVS, and highly accurate prediction is constructed to screen promising phosphorescent Pt(II) emitters. Based on the core structures and the functional groups, more than 3600 synthesis-friendly complexes are generated. A feature strategy concerning core structures, molecular properties and testing medium is established to represent each complex correctly and validly. To rapidly screen candidate molecules, different ML algorithms are utilized and compared based on a dataset of 198 Pt-based emitters. The feature importance analysis reveals that molecular properties such as molecular weight of heavy atoms and the fraction of C atoms that are sp^3 hybridized, similarity fingerprints like FpDensityMorgan, and topological descriptors such as Kappa show significant contributions in the predictions. Recently reported Pt-complexes are introduced as external samples to evaluate the capability of the HTVS-ML models, which indicates the generality of the optimal models. Under the favor of a three-tier well-designed screening rule and highly accurate predictions approach, 12 most promising complexes are ultimately selected and recommended for further synthesis. Importantly, two of them are very similar to the emitters with excellent performance. This work presents the first ML-based general protocol for generating, HTVS and accurately evaluating vital photophysical properties of Pt-emitters. We expect the protocol will be beneficial in rationally designing Pt-emitters and thus help discover novel OLED materials with distinguished performances.

5. Future perspective

In the future outlook, significant advancements are expected through the integration of artificial intelligence (AI), ML, and quantum chemistry. The potential for transformative breakthroughs lies in the convergence of these technologies, focusing on themes such as materials property prediction, screening, and inverse design. Additionally, accelerated materials simulations and explainable ML methods will play a crucial role in shaping the future landscape.

Within the realm of developing Pt(II) complexes for OLED applications, the integration of HTVS-ML and Δ-learning stands out as a highly promising approach. The proposed robust protocol not only streamlines complex generation and HTVS but also delivers accurate predictions of key photophysical properties. This approach reduces the experimental workload and instils confidence in the rational design and rapid development of novel OLED materials with tailored properties.

Looking ahead, the continuous progress in AI, ML, and quantum chemistry, particularly Δ-learning in materials science, holds immense promise for expediting materials discovery and optimization across various applications, including OLED technology. Future endeavours will focus on addressing the challenges such as efficient representations of complex materials and establishing advanced frameworks for predicting properties of low-sample materials, where experimental and computational data are scarce.

Data and code availability

All data and code employed in this work are available from an open GitHub repository: https://github.com/JerryShuaiWANG/Pt-HTVS-ML.

Acknowledgment

This work is financially supported by the RGC General Research Fund under Grant No. 17309620, Hong Kong Quantum AI Lab Limited and Air @ InnoHK of Hong Kong Government. C Y Y acknowledges the support from National Natural Science Foundation of China (Grant Nos. 22073007 and 22473022) and Shenzhen Basic Research Key Project Fund (Grant No. JCYJ20220818103200001). L H H thanks financial support from National Natural Science Foundation of China (No. 22273010) and Department of Science and Technology of Jilin Province (20210402075GH). C-M C and F-F H acknowledge Guangdong Major Project of Basic and Applied Basic Research (Grant No. 2019B030302009).

Author contributions and details

Shuai Wang: writing—original draft, investigation, software, data curation, formal analysis. ChiYung Yam: supervision, conceptualization, writing—review & editing. LiHong Hu: conceptualization, methodology, writing—review & editing. Faan-Fung Hung: data curation, resources, writing—review & editing. Shuguang Chen: project administration, resources. Chi-Ming Che: supervision, conceptualization, funding acquisition. GuanHua Chen: supervision, conceptualization, funding acquisition, writing—review & editing.

Conflict of interest

There are no conflicts to declare.

ORCID iDs

Shuai Wang https://orcid.org/0000-0002-3195-2691
ChiYung Yam https://orcid.org/0000-0002-3860-2934
LiHong Hu https://orcid.org/0000-0003-3792-2917
Faan-Fung Hung https://orcid.org/0009-0001-3634-1355
Chi-Ming Che https://orcid.org/0000-0002-2554-7219

References

[1] Saeki A and Kranthiraja K 2019 A high throughput molecular screening for organic electronics via machine learning: present status and perspective *Jpn. J. Appl. Phys.* **59** SD0801
[2] Paterson L, May F and Andrienko D 2020 Computer aided design of stable and efficient OLEDs *J. Appl. Phys.* **128** 160901
[3] Hong G, Gan X, Leonhardt C, Zhang Z, Seibert J, Busch J M and Bräse S 2021 A brief history of OLEDs—emitter development and industry milestones *Adv. Mater.* **33** 2005630
[4] Li G and She Y 2018 Tetradentate cyclometalated platinum(II) complexes for efficient and stable organic light-emitting diodes *Light-Emitting Diode—An Outlook on the Empirical Features and Its Recent Technological Advancements* (intechopen)
[5] Li K, Tong G S M, Wan Q, Cheng G, Tong W-Y, Ang W-H, Kwong W-L and Che C-M 2016 Highly phosphorescent platinum(II) emitters: photophysics, materials and biological applications *Chem. Sci.* **7** 1653–73
[6] Li H, Lam T-L, Tan X, Dai L and Che C-M 2021 26-3: invited paper: high performance and long device lifetime organic light-emitting diodes using a tetradentate platinum (II) emitter *SID Symp. Dig. Tech. Pap.* **52** 328–31
[7] Pander P, Daniels R, Zaytsev A, Horn A, Sil A, Penfold T, Williams G, Kozhevnikov V and Dias F 2021 Exceptionally fast radiative decay of a dinuclear platinum complex through thermally activated delayed fluorescence *Chem. Sci.* **12** 6172–80
[8] Li H, Lam T-L, Yan L, Dai L, Choi B, Cho Y-S, Kwak Y and Che C-M 2020 Tetradentate platinum(II) emitters: design strategies, photophysics, and OLED applications *Liquid Crystals and Display Technology* (intechopen)
[9] Parr R G and Yang W 1989 *Density-functional Theory of Atoms and Molecules* (Oxford University)
[10] Marques M, Rubio A, Gross E K, Burke K, Nogueira F and Ullrich C A 2006 *Time-dependent Density Functional Theory* (Springer)
[11] Brlec K, Savory C N and Scanlon D O 2023 Understanding the electronic structure of $Y_2Ti_2O_5S_2$ for green hydrogen production: a hybrid-DFT and GW study *J. Mater. Chem.* A **11** 16776–87
[12] Dixit A, Annie Abraham J, Manzoor M, Altaf M, Anil Kumar Y and Sharma R 2024 A comprehensive DFT analysis of the physical, optoelectronic and thermoelectric attributes of Ba_2InNbO_6 double perovskites for eco-friendly technologies *Mater. Sci. Eng.* **307** 117530
[13] Koval A M, Jenness G R, Schutt T C, Kosgei G K, Fernando P U A I and Shukla M K 2024 Periodic DFT calculations to compute the attributes of a quantum material: honeycomb ruthenium trichloride *Phys. Chem. Chem. Phys.* **26** 19369–79
[14] Besley N A 2021 Modeling of the spectroscopy of core electrons with density functional theory *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **11** e1527
[15] Glossman-Mitnik D 2022 *Density Functional Theory: Recent Advances, New Perspectives and Applications* (intechopen)
[16] González L, Escudero D and Serrano-Andrés L 2012 Progress and challenges in the calculation of electronic excited states *ChemPhysChem* **13** 28–51
[17] DiLuzio S, Mdluli V, Connell T U, Lewis J, VanBenschoten V and Bernhard S 2021 High-throughput screening and automated data-driven analysis of the triplet photophysical properties of structurally diverse, heteroleptic iridium(III) complexes *J. Am. Chem. Soc.* **143** 1179–94
[18] Curtarolo S, Hart G L W, Nardelli M B, Mingo N, Sanvito S and Levy O 2013 The high-throughput highway to computational materials design *Nat. Mater.* **12** 191–201
[19] Shoichet B K 2004 Virtual screening of chemical libraries *Nature* **432** 862–5
[20] Boonpalit K, Chuntakaruk H, Kinchagawat J, Wolschann P, Hannongbua S, Rungrotmongkol T and Nutanong S 2024 Pre-training strategy for antiviral drug screening with low-data graph neural network: a case study in HIV-1 K103N reverse transcriptase *J. Comput. Chem.* **46** e27514
[21] Krasoulis A, Antonopoulos N, Pitsikalis V and Theodorakis S 2022 DENVIS: scalable and high-throughput virtual screening using graph neural networks with atomic and surface protein pocket features *J. Chem. Inf. Model.* **62** 4642–59
[22] Damm-Ganamet K L *et al* 2019 Accelerating lead identification by high throughput virtual screening: prospective case studies from the pharmaceutical industry *J. Chem. Inf. Model.* **59** 2046–62
[23] Parvatikar P P, Patil S, Khaparkhuntikar K, Patil S, Singh P K, Sahana R, Kulkarni R V and Raghu A V 2023 Artificial intelligence: machine learning approach for screening large database and drug discovery *Antiviral Res.* **220** 105740
[24] Tripathi N M and Bandyopadhyay A 2022 High throughput virtual screening (HTVS) of peptide library: technological advancement in ligand discovery *Eur. J. Med. Chem.* **243** 114766
[25] Cui L, Li Q, Zhang Y, Zhang J, Wang Z, Chen J and Zheng B 2024 Machine learning-assisted high-throughput screening of transparent organic light-emitting diode anode materials *Chem. Sci.* **15** 19375–89
[26] Omar Ö H, Del Cueto M, Nematiaram T and Troisi A 2021 High-throughput virtual screening for organic electronics: a comparative study of alternative strategies *J. Mater. Chem.* C **9** 13557–83
[27] Pyzer-Knapp E O, Suh C, Gómez-Bombarelli R, Aguilera-Iparraguirre J and Aspuru-Guzik A 2015 What is high-throughput virtual screening? A perspective from organic materials discovery *Annu. Rev. Mater. Res.* **45** 195–216
[28] Kim J, Mok D H, Kim H and Back S 2023 Accelerating the search for new solid electrolytes: exploring vast chemical space with machine learning-enabled computational calculations *ACS Appl. Mater. Interfaces* **15** 52427–35
[29] Mao Y, Yao X, Yu Z, An Z and Ma H 2024 Ground-state orbital descriptors for accelerated development of organic room-temperature phosphorescent materials *Angew. Chem., Int. Ed.* **63** e202318836
[30] Quirós M, Gražulis S, Girdzijauskaitė S, Merkys A and Vaitkus A 2018 Using SMILES strings for the description of chemical connectivity in the crystallography open database *J. Cheminf.* **10** 23
[31] Huang D and Cole J M 2024 A database of thermally activated delayed fluorescent molecules auto-generated from

scientific literature with ChemDataExtractor *Sci. Data* **11** 80
[32] Lopez S A, Pyzer-Knapp E, Simm G, Lutzow T, Li K, Seress L R, Hachmann J and Aspuru-Guzik A 2016 The Harvard organic photovoltaic dataset *Sci. Data* **3** 2052–4463
[33] Joung J F, Han M, Jeong M and Park S 2020 Experimental database of optical properties of organic compounds *Sci. Data* **7** 295
[34] Ramakrishnan R, Dral P O, Rupp M and von Lilienfeld O A 2014 Quantum chemistry structures and properties of 134 kilo molecules *Sci. Data* **1** 140022
[35] Omar Ö H, Nematiaram T, Troisi A and Padula D 2022 Organic materials repurposing, a data set for theoretical predictions of new applications for existing compounds *Sci. Data* **9** 54
[36] Balcells D and Skjelstad B B 2020 tmQM dataset—quantum geometries and properties of 86k transition metal complexes *J. Chem. Inf. Model.* **60** 6135–46
[37] 2023 Schrödinger Release 2023-2: maestro (Schrödinger, LLC)
[38] RDKit: Open-Source Cheminformatics Software (available at: www.rdkit.org/) (Accessed 17 February 2023)
[39] Wang S, Yam C, Chen S, Hu L, Li L, Hung F-F, Fan J, Che C-M and Chen G 2024 Predictions of photophysical properties of phosphorescent platinum(II) complexes based on ensemble machine learning approach *J. Comput. Chem.* **45** 321–30
[40] Baerends E J, Ziegler T, Autschbach J, Bashford D, Bérces A, Bickelhaupt F M, Bo C, Boerrigter P M and Cavallo L 2021 ADF2021, SCM
[41] Kennard R W and Stone L A 1969 Computer aided design of experiments *Technometrics* **11** 137–48
[42] Galvão R K H, Araujo M C U, José G E, Pontes M J C, Silva E C and Saldanha T C B 2005 A method for calibration and validation subset partitioning *Talanta* **67** 736–40
[43] Todeschini R and Consonni V 2008 *Handbook of Molecular Descriptors* (Wiley)
[44] Hall L H and Kier L B 1991 The Molecular connectivity Chi indexes and Kappa shape indexes in structure-property modeling *Reviews in Computational Chemistry* (Wiley) pp 367–422
[45] Breiman L 2001 Random forests *Mach. Learn.* **45** 5–32
[46] Palit I and Reddy C K 2012 Scalable and parallel boosting with MapReduce *IEEE Trans. Knowl. Data Eng.* **24** 1904–16
[47] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q and Liu T-Y 2017 LightGBM: a highly efficient gradient boosting decision tree *Advances in Neural Information Processing Systems*
[48] Chen T and Guestrin C 2016 XGBoost: a scalable tree boosting system (New York) *Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining* pp 785–94
[49] Cortes C and Vapnik V 1995 Support-vector networks *Mach. Learn.* **20** 273–97
[50] Cover T and Hart P 1967 Nearest neighbor pattern classification *IEEE Trans. Inf. Theory* **13** 21–27
[51] Murphy K P 2012 *Machine Learning: A Probabilistic Perspective* (MIT Press)
[52] Bergstra J, Komer B, Eliasmith C, Yamins D and Cox D D 2015 Hyperopt: a Python library for model selection and hyperparameter optimization *Comput. Sci. Discov.* **8** 014008
[53] Riniker S and Landrum G A 2013 Similarity maps—a visualization strategy for molecular fingerprints and machine-learning methods *J. Cheminf.* **5** 43
[54] Balaban A T 1982 Highly discriminating distance-based topological index *Chem. Phys. Lett.* **89** 399–404
[55] Labute P 2000 A widely applicable set of descriptors *J. Mol. Graph. Model.* **18** 464–77
[56] Ertl P, Rohde B and Selzer P 2000 Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties *J. Med. Chem.* **43** 3714–7
[57] Lam T-L, Li H, Tan K, Chen Z, Tang Y-K, Yang J, Cheng G, Dai L and Che C-M 2024 Sterically hindered tetradentate [Pt(O^N^C^N)] emitters with radiative decay rates up to 5.3×10^5 s^{-1} for phosphorescent organic light-emitting diodes with LT_{95} lifetime over 9200 h at 1000 cd m^{-2} *Small* **20** 2307393