Combined Machine Learning, Computational and Experimental Analysis of the Iridium(III) Complexes with Red to Near-IR Emission

*Anas Karuth[1], Gerardo M. Casanola-Martin[1], Levi Lystrom[2], Wenfang Sun[2], Dmitri Kilin[2], Svetlana Kilina[2], Bakhtiyor Rasulev[1]**

[1]Coatings and Polymeric Materials, North Dakota State University, Fargo, ND 58108, United States

[2]Department of Chemistry and Biochemistry, North Dakota State University, Fargo, ND 58108, United States

*Corresponding author: e-mail: bakhtiyor.rasulev@ndsu.edu

Keywords: iridium(III) complexes, near-IR emission, computational, machine learning, QSPR, predictive modeling

Abstract

Various coordination complexes have been the subject of experimental or theoretical studies in recent decades because of their fascinating photophysical properties. In this work a combined experimental and computational approach applied to investigate the optical properties of monocationic Ir(III) complexes. In result, an interpretative machine learning-based Quantitative Structure-Activity Relationship (QSAR) model was successfully developed, which can reliably predict the emission wavelength of the Ir(III) complexes and provides foundations for theoretical

evaluation of the optical properties of Ir(III) complexes. A hypothesis was proposed to mechanistically explain the differences in emission wavelengths between structurally different individual Ir(III) complexes. To the best of our knowledge, this is the first attempt to develop predictive machine learning (QSAR) model for the optical properties of Ir(III) complexes. The efficacy of the developed model was demonstrated by high R^2 values for the training and test sets of 0.84 and 0.87, respectively, and by performing the validation using y-scrambling techniques. A notable relationship between the N-N distance in the diimine ligands of the Ir(III) complexes and emission wavelengths was revealed. This combined experimental and computational approach shows a great potential for rational design of new Ir(III) complexes with desired optical properties. Moreover, the developed methodology could be extended to other octahedral transition-metal complexes.

Keywords: iridium(III) complexes, emission, ligands, light absorption, substituents

1. INTRODUCTION

In recent decades, various transition-metal coordination complexes have been the subject of extensive studies because of their fascinating photophysical properties. Among these transition-metal complexes, Ir(III) complexes are particularly interesting due to their strong spin-orbit coupling, which gives rise to an efficient intersystem crossing and high phosphorescence quantum yield at room temperature.[1-3] Because of these characteristics, Ir(III) complexes have been widely studied for organic light-emitting devices (OLEDs),[4, 5] light-emitting electrochemical cells (LEECs)[6, 7] luminescent biological-labeling reagents[8], chemosensors[9], and upconversion applications.[10] In addition, it has been found that Ir(III) complexes can exhibit efficient reverse

saturable absorption (RSA).[11-14] RSA is a nonlinear optical phenomenon in which the materials exhibit a standard linear absorption at low incident laser pulse, while their absorptivity increases when the incident laser power increases. RSA is a required property for photonic applications like optical switching, optical rectification, laser pulse shaping, and compression.[15, 16] To produce efficient RSA with nanosecond laser pulses, the lowest triplet state should have a long lifetime, a strong absorption coefficient, and a high quantum yield. For practical applications, the broadband RSA materials are also required to cover the spectral range of 500-900 nm.

To reach this goal, various structural modifications of Ir(III) complexes were suggested to tune both the ground- and excited-state properties and shift absorption and phosphorescence of these complexes to the near infrared (NIR) spectral range, while also improve RSA. For instance, the optical properties of Ir(III) complexes can be altered by using different types of diimine or cyclometalating ligands or varying substituents on the ligands.[12, 17, 18] The previous studies show that the emission properties of Ir(III) complexes consisting cyclometalated ligands 2-phenylpyridine (ppy, C-N ligand) and ancillary ligands 2,2'-bipyridine (bpy, N-N ligand) can be controlled by substitution of electron-withdrawing and donating groups.[13, 17] Extensive electrochemical and computational studies show that the lowest unoccupied molecular orbital (LUMO) of these Ir(III) complexes is typically located on the N^N ligand; while the highest occupied molecular orbital (HOMO) is delocalized on the C^N ligands and the d-orbital of the Ir(III) center.[11] Thus, electron-donating or withdrawing substituents on the N^N ligand could lead to hypsochromic or bathochromic shifts of the lowest-energy charge transfer optical transition by destabilization or stabilization of the excited electron state, while the hole state is negligibly effected.[19]

Beside the electron withdrawing/donating groups, introducing π-conjugated substituents on the N^N ligand could alter the nature of the lowest triplet excited state. The most frequently used π–conjugated N^N ligand is 2-arylimidazole[4,5-f][1,10]phenanthroline is widely used in biomolecular targeting and chemosensing.[20-22] Even though these complexes exhibit good photophysical properties, its sensitivity to environmental changes impacts its stability during the fabrication and application. To overcome this issue, Sun et al. studied a series of 2-aryloxazolo[4,5-f][1,10]-phenanthroline ligands (N^N ligands) and their cationic Ir(III) complexes with varying electron withdrawing or donating groups and π–conjugation length.[13, 23] The spectroscopic and quantum-mechanical study reveal increased co-planarity with 2-pehenyl substitution in N^N ligands, which promotes increased emission intensity in the solid and solution state of Ir(III) complexes. It was also shown that the π–extended substituents at the 2-position of the oxazolo[4,5-f][1,10]phenanthroline ligand show stronger RSA than other complexes with phenyl or substituted phenyl substitution in N^N ligands.[23]

Extending π–conjugation can be also achieved via benzannulation of N^N or C^N ligands of Ir(III) complexes.[11] Depending on the site of benzannulation in C^N ligands, the Ir(III) complexes can exhibit a blue or red shift of the lowest-energy absorption and emission bands. This is attributed to the site-dependent destabilization or stabilization of the excited electron state upon benzannulation in the Ir(III) complexes. It was shown that incorporation of quinoxaline or benzo[*g*]quinoxaline unit into either the N^N or C^N ligand red−shifted the spin−forbidden $^3\pi,\pi^*/^3$CT (charge transfer) absorption bands into longer wavelengths, while reducing the triplet excited−state lifetimes.[24-26] [21,31,27] In contrast, Ir(III) bipyridine complex bearing 1,2–diphenyl–9*H*–pyreno[4,5–*d*]imidazole (dppi)[27] [28] or naphthalimidyl (NI)[28] [24] cyclometallating ligands have displayed a long−lived strong triplet excited−state absorption in the spectral regions of 410 − 800

nm. The synergetic role of dppi C^N ligands and pyrazine–based N^N ligands in the red−shift of the ground−state absorption to the far-red/NIR regions while keeping a relatively long−lived and broadly absorbing triplet excited state have been recently demonstrated by Sun group, and also showing high RSA of these complexes at 532 nm for ns laser pulses.[14]

Despite these abundant experimental and computational studies on the non-linear and linear optical properties of Ir(III) complexes,[13, 23-36 20–33] establishing the relationship between the structure of ligands and emission properties of Ir(III) complexes for their rational design is still illusive task, due to many contradictive and contra-intuitive data. In this regard, machine learning methods can be very helpful to find the structure-property correlations in Ir(III) complexes, which can combine various data and extract the essential knowledge. Thus, the data driven modeling techniques like quantitative structure-activity/property relationship (QSAR/QSPR) can be used to establish mathematical correlations between optical features of Ir(III) complexes and various structural factors like the π–conjugation of different type of ligands (N^N ligands and C^N ligands), benzannulation on different ligand's sites, electron withdrawing/donating substituents, and beyond. QSAR modeling has been successfully used to model structure and the ground state electronic property models for polymers, organic and inorganic materials, semiconductors, and organometallic complexes.[37-39] QSAR has been successfully applied to generate a toxicity model of organometallic complexes.[40] In another work, the relative binding affinity of diverse organometallic estradiol derivatives is also modeled using the QSAR approach.[41] However, no QSAR studies have been reported correlating the phosphorescence properties of Ir(III) complexes and the structure of complexes/ligands. Deriving a mathematical relationship between the emission or absorption wavelength of Ir(III) complexes and structural features of ligands in the Ir(III)

complexes obtained from QSAR/QSPR models can help to synthesize the Ir(III) complexes with tailormade optical performance and characteristics.

In this work, we have collected 46 different Ir(III) complexes with diverse structural features that were synthesized and characterized experimentally, as reported in literature,[13, 17, 23-36] to find a relationship between near-IR emission property and the structure, and then develop a quantitative structure-property relationship (QSPR) model. The investigated Ir(III) complexes have a diverse degree of substitution, including electron withdrawing and donating groups, varying π-conjugation length on C^N and/or N^N ligands, and benzannulation on different ligand's sites. This work is envisioned to predict the triplet emission wavelength of a set of Ir(III) complexes with structural and morphological complexities by using the unique combination of experiment, quantum chemistry, and machine learning-based QSAR methods, as illustrated in **Fig. 1**. An interpretative machine learning-based QSAR model was successfully developed, which can reliably predict the emission wavelength of the Ir(III) complexes based on their structural modifications and provide foundations for theoretical evaluation of the optical properties of the Ir(III) complexes. It is worth noting that unexpected relation between a decrease in the N-N distance of a pair of nitrogens coordinated to the metal center of the Ir(III) complexes and the redshift in the emission wavelengths is revealed. Interestingly, this relationship has never been discussed in literature before. We associate these correlations to a degree of distortion in the octahedral structure of complexes resulting in the perturbed ligand field. Overall, our combined experimental and QSAR-based computational approach opens a pathway for cost-effective and efficient development of optical materials for various applications.

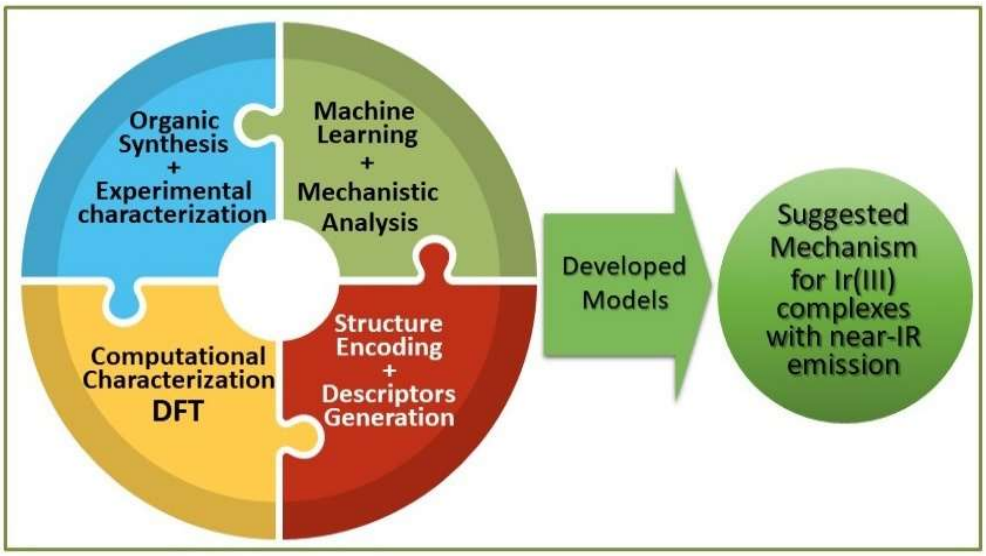

Figure 1. Scheme of the workflow applied to investigate structurally different iridium (III) complexes by combined experimental, computational and machine learning methods.

2. MATERIALS AND METHODS

The sequence of the tasks in this work to develop a QSPR model is based on following steps: (1) data set selection; (2) molecular structure generation; (3) geometry optimization of molecular structures using appropriate methods; (4) a set of various descriptors generation; (5) variable selection and/or data reduction methods; (6) model generation; (7) and validation and predictability evaluation of the developed models.[42, 43 34,35]

2.1. Data Set Selection: Iridium(III) complexes emission data set

The experimental values of the *emission wavelengths* of 46 different Ir(III) complexes are collected from various sources. [13, 17, 23-36] The data set of these complexes is split into training and test sets (80% of 46 complexes as training and 20% as a test set). The splitting of the data set into training

and test set was carried out by arranging the data set in ascending order and taking every 5th chemical from the data set as a test or prediction set.

Table 1. Dataset of investigated 46 organometallic compounds

ID	Status	Experimental $\lambda_{emission}$, nm	Predicted $\lambda_{emission}$, nm	Δ $\lambda_{emission}$	Reference
1	Training	567	582.91	15.91	34
2	Training	570	615.48	45.48	27
3	Training	577	582.34	5.34	13
4	Training	578	580.07	2.07	34
5	Prediction	579	578.37	-0.63	13
6	Training	580	582.34	2.34	13
7	Training	580	601.87	21.87	27
8	Training	582	606.97	24.97	36
9	Training	582	582.34	0.34	13
10	Prediction	584	581.77	-2.23	13
11	Training	584	582.34	-1.66	13
12	Training	584	614.09	30.09	34
13	Training	584	614.09	30.09	33
14	Training	585	582.34	-2.66	13
15	Prediction	585	601.87	16.87	27
16	Training	586	582.34	-3.66	13
17	Training	587	567.03	-19.97	13
18	Training	587	601.87	14.87	27
19	Training	592	567.03	-24.97	13
20	Excluded	593			23
21	Training	593	600.03	7.03	34
22	Training	595	581.77	-13.23	13
23	Training	600	606.97	6.97	36
24	Training	600	614.66	14.66	34
25	Prediction	601	604.00	3.00	32
26	Training	602	601.37	-0.63	32
27	Training	604	601.87	-2.13	27
28	Training	608	604.00	-4.00	36
29	Training	608	640.85	32.85	17
30	Prediction	609	604.56	-4.44	36
31	Training	612	573.38	-38.62	32
32	Training	619	606.40	-12.60	33
33	Training	625	668.21	43.21	23
34	Training	630	638.16	8.16	33
35	Prediction	645	714.70	69.70	23

36	Training	656	638.72	-17.28	33
37	Training	657	675.01	18.01	17
38	Training	657	679.55	22.55	23
39	Training	658	640.85	-17.15	17
40	Prediction	672	638.16	-33.84	33
41	Training	680	601.87	-78.13	27
42	Training	755	710.73	-44.27	23
43	Training	780	714.14	-65.86	23
44	Training	794	765.62	-28.38	30
45	Prediction	794	817.79	23.79	30
46	Training	801	801.25	0.25	30
47	Training	801	829.13	28.13	30

2.2. Molecular Structure Generation: DFT structure optimization

All complexes of this series were optimized at the density functional (DFT) level of theory using Gaussian-09/16 software package.[44] The geometries of complexes were optimized in their singlet and triplet spin configurations by DFT using the hybrid Perdew, Burke, and Ernzerh of functional (PBE1PBE)[45, 46] and the LANL2DZ[47] basis set assigned for Ir(III) and 6-31G* basis set for all the remaining atoms.[48, 49] Dichloromethane was used as the solvent within the conductor-like polarizable continuum model (CPCM)[50] reaction field method. This methodology is described in detail in our previous publications and has been shown to provide good agreement with experimental structural and optical data of reported Ir(III) complexes. [13, 17, 23-35]

2.3. Descriptors Generation

The molecular descriptors were calculated for all complexes in this work, where descriptors are mathematical representations of chemical information contained in a molecule. The chemical structure is built using ChemSketch software,[51] then computed and optimized by DFT methods and then optimized structure used for descriptors' degeneration by Dragon 6 software.[52] The Dragon 6 software generates more than 4500 various descriptors corresponding to 0D-, 1D-, 2D-,

and 3D-structure-based indexes. The descriptors are comprised of 20 different classes — constitutional, topological, walk and path counts, connectivity indices, information indices, 2D autocorrelations, edge adjacency indices, Burden eigenvalues, topological charge indices, eigenvalue based indices, Randic molecular profiles, geometrical descriptors, RDF descriptors, 3D-MoRSE descriptors, WHIM descriptors, GETAWAY descriptors, functional groups, atom-centered fragments, charge descriptors, and molecular property descriptors.[53] After filtering constant and near-constant descriptors (>0.95 similarity), about 2863 descriptors were generated per iridium organometallic structure.

2.4. QSAR Model Generation and Validation

Machine Learning-based QSAR is an approach that searching for a mathematical relationship between a biological activity or physical property of a molecular system and its geometric, chemical, or physical characteristics. QSAR attempts to find a consistent relationship between a biological activity or physical property and molecular properties so that these "rules" can be used to evaluate the activity of new chemical systems. Once a valid QSAR has been determined, it should be possible to predict the physical property or biological activity of related compounds before they are put through expensive and time-consuming biological or physical testing.

In this work, the correlation between near-IR emission activity and structural properties was developed by using the approach which consists of a Genetic Algorithm (GA)[54] for variable selection and Multiple Linear Regression Analysis (MLRA) for final mathematical model generation. Thus, preliminary model selection was performed using the GA-MLRA[43] technique as implemented in the QSARINS 2.2 program.[55] It is worth noting that genetic algorithms have been applied in recent studies as a powerful tool to address many problems in QSAR studies.[37, 43, 56]

The GA algorithm was used as a variable selection technique to reduce the number of descriptors in a final model. The GA variable selection procedure started with a population of 500 random models and 5000 iterations to evolve with the mutation probability specified at 85%. The systematic study delineated here consists of several QSAR models (1 to 7 variables in each model), followed by statistical analysis with evaluation by a squared correlation coefficient R^2, root means square error (RMSE), Fisher coefficient (F) to find the best performing model.

To ensure the reliability of data produced by models, the validation of the QSPR model is very important. Thus, internal and external validations are considered to be necessary for checking the robustness of the model. In the internal validation process, the final set of QSAR models was generated by using the GA-MLR approach and then tested by applying the cross-validation "leave-one-out" technique (the process of removing a molecule from the set, then creating and validating the model against the individual molecules, which was performed for the entire training set) Q^2_{LOO}. Thus, we utilized the following equations to calculate the correlation coefficient, R^2 (Eq. 1) and the root mean square error of calibration (training) $RMSE_C$, as the measures of goodness-of-fit for each developed model (Eq. 2).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i^{obs} - y_i^{pred})^2}{\sum_{i=1}^{n}(y_i^{obs} - \tilde{y}^{obs})^2} \quad (1)$$

$$RMSE_C = \sqrt{\frac{\sum_{i=1}^{n}(y_i^{obs} - y_i^{pred})^2}{n}} \quad (2)$$

In leave-one-out (LOO) method of cross validation (*cv*), the process of removal of a molecule and, creating and validating against the individual molecules is performed for the entire training set. Once completed for all the mean Q^2 value is reported. Cross-validation (*CV*), is utilized to measure the predictive ability of the model and to avoid over-fitting scenarios.

We calculated the cross-validated coefficient Q^2_{LOO} (leave-one-out method) and root mean square error of cross-validation $RMSE_{CV}$. Both statistics were calculated according to Eq. 3 and Eq.4:

$$Q^2_{LOO} = 1 - \frac{\sum_{i=1}^{n}(y_i^{obs} - y_i^{predcv})^2}{\sum_{j=1}^{n}(y_j^{obs} - \tilde{y}^{obs})^2} \quad (3)$$

$$RMSE_{CV} = \sqrt{\frac{\sum_{i=1}^{n}(y_i^{obs} - y_i^{predcv})^2}{n}} \quad (4)$$

In the process of model validation, external validation is indispensable. The external validation is checked for its ability to predict new compounds which are not included in the training set. This is done by applying the model equation obtained from the training set to the test data set. The external predictivity of each model is represented as R^2_{Test} (Eq.5) and root mean square error of prediction ($RMSE_P$) (Eq.6). In this work, we applied all these three statistics (R^2, Q^2_{LOO}, R^2_{Test})

$$R^2_{Test} = 1 - \sum_{j=1}^{k}(y_j^{obs} - y_j^{pred})^2 / \sum_{j=1}^{k}(y_j^{obs} - \hat{y}^{obs})^2 \quad (5)$$

$$RMSE_{Test} = \sqrt{\frac{\sum_{i=1}^{k}(y_j^{obs} - y_j^{pred})^2}{k}} \quad (6)$$

Where: y_j^{obs} – the experimental (observed) value of the property for the i^{th}/j^{th} compound; $y_j{}^{pred}$ – predicted value for i^{th}/j^{th} compound. $\tilde{y}$, $\hat{y}$ – the mean experimental value of the property in the training and validation set, respectively; n, k – the number of compounds in the training and validation set, respectively.

Additionally, the chemical applicability domain (AD) for the models obtained was calculated by the leverage approach to verify predictive reliability.[57] To visualize the applicability

domain of the QSPR models, the Williams plot was used. Thus, Williams's plot of standardized cross-validated residuals (RES) vs. leverage (Hat diagonal) values (HAT) clearly depicts both the response outliers (Y outliers) and structurally influential compounds (X outliers) in a model.

3. RESULTS AND DISCUSSION

In this work, we have investigated 46 Ir(III) complexes with diverse structural features that were synthesized and characterized experimentally and computationally, to find a relationship between near-IR emission values and the structure and develop a QSPR model.

The model building is done systematically from the first model to the seventh model by accessing the regression coefficient of the training set in each model and followed by validation of the test set (**Table 1**).

Table 1. Shows the list of models (1-7 model) and associated descriptors for each model.

Model	Descriptors	$R^2_{Training}$	$RMSE_C$	Q^2_{LOO}	$CCC_{Training}$	F-Test	R^2_{Test}	$RMSE_{Test}$
Model 1	GATS4i	0.55	0.10	0.5	0.51	40.50	0.67	0.08
Model 2	SpMin3_Bh(m) F08[N-N]	0.81	0.09	0.74	0.62	31.00	0.75	0.07
Model 3	SpMin3_Bh(m) Eig04_EA(dm) B08[N-N]	0.84	0.08	0.78	0.71	31.09	0.87	0.06
Model 4	SpMin3_Bh(m) Chi0_EA(dm) Mor29s F08[N-N]	0.89	0.08	0.84	0.79	34.54	0.84	0.06
Model 5	GATS5i SpMin3_Bh(m) Mor28i Mor29s F08[N-N]	0.92	0.07	0.86	0.81	32.10	0.80	0.06
Model 6	SM1_Dz(i) GATS5i SpMin3_Bh(m) Mor28i Mor29s F08[N-N]	0.93	0.07	0.87	0.85	33.20	0.77	0.06
Model 7	ZM2Kup GATS5i SpMin3_Bh(m) Mor28m Mor21s Mor29s F08[N-N	0.95	0.06	0.91	0.86	30.80	0.68	0.06

It can be observed from the **Table 1** that the $R^2_{Training}$ values show a marching trend with an increase of the number of variables in the model and the same behavior is exhibited by R^2_{Test} but up to 3 variables only and then showed plummeting behavior which manifests that the predictive capability of the model beyond the 3 variables is decreased due to overfitting issues. At the same time the predictive coefficients (R^2 and Q^2) of the training set increases for higher magnitude models (4 to 7), while its predictive capability for the test set declines. This trend is illustrated in **Fig. 2**. Overall, the 3-variable model shows a good combination of high $R^2_{Training}$ and $R^2_{Test/validation}$ values and, therefore, is selected as the best and most accurate model among all 1 to 7 variable models.

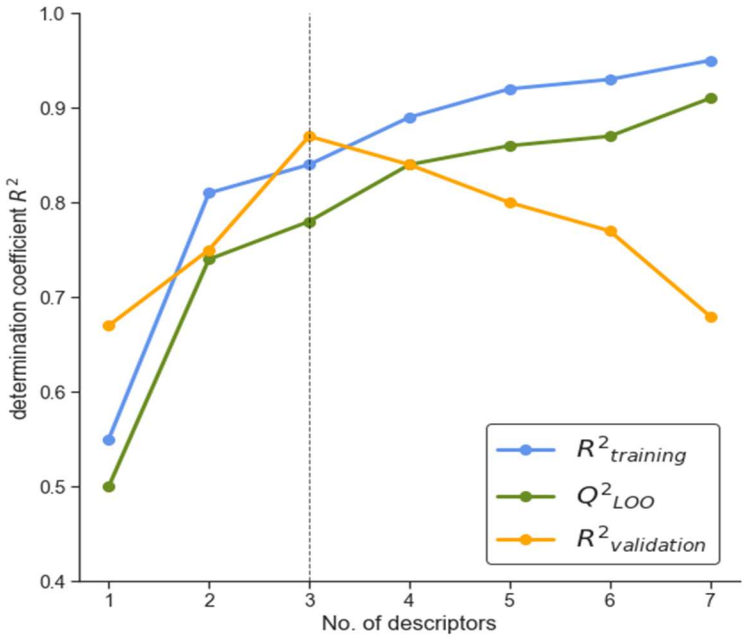

Figure 2. Statistical data on R^2 and Q^2 for models 1-7 for training and R^2 for external (test) set.

The selected 3-variable model predicts the *emission wavelength* (nm) according to the following equation (**Eq. 7**):

$$\boldsymbol{\lambda_{emission}} = 567.03*\boldsymbol{SpMin3\_Bh(m)} - 50.23*\boldsymbol{Eig04\_EA(dm)} + 149.5*\boldsymbol{B08[N\text{-}N]} - 537.26, \quad (\mathbf{7})$$

$n = 46, \quad R^2_{train} = 0.84, \quad R^2_{test} = 0.87, \quad Q^2_{loo} = 0.78, \quad RMSE_{train} = 26.72,$

$RMSE_{cv} = 31.2, \quad RMSE_{test} = 29.3, \quad F = 61.3$

where the descriptors are the specific structure-related indexes and represented in **Table 2**, while obtained statistical parameters are explained in Materials and Methods section and Table 2.

Table 2. The list of descriptors for 3-variable model, their type and short description.

Descriptor	Descriptor information	Type
SpMin3_Bh(m)	smallest eigenvalue n. 3 of Burden matrix weighted by mass	Burden eigenvalues descriptors
Eig04_EA(dm)	eigenvalue n. 4 from edge adjacency mat. weighted by dipole moment	Edge adjacency indices
B08[N-N]	Presence/absence of N - N at topological distance 8	2D atom pair

The descriptors of 3-variable model shown in **Table 2** can be broadly classified as a 2D-matrix, 3D-MoRSE, gateway, functional, atom pair, and electrotopological indices descriptors. The model shows that the emission wavelength (nm) strongly correlates with such molecular descriptors as weighted mass, weighted dipole moment, and N-N topological distance in the ligands of investigated Ir(III) complexes. Moreover, the mass or bulkiness of the ligand significantly influences the emission wavelength, thereby contributing to the optical properties of complexes.

Figure 3 shows the magnitude of the influence of various descriptors on the emission wavelength value (λ, nm) according to the 3-variable model (**Eq. 7**). The positive or negative influence of each descriptor on λ is represented with corresponding coefficient value. The SpMin3_Bh(m) and B08[N-N] show a positive impact on the emission wavelength value (λ, nm) of the iridium (III)

complexes, indicating the red shift of wavelength (λ, nm). The model shows that larger mass of the ligand and larger topological distance between N-N atoms in the ligand can red shift the emission wavelength, on other hand the as the larger dipole moment of the ligand tends to reduce the emission wavelength (λ, nm).

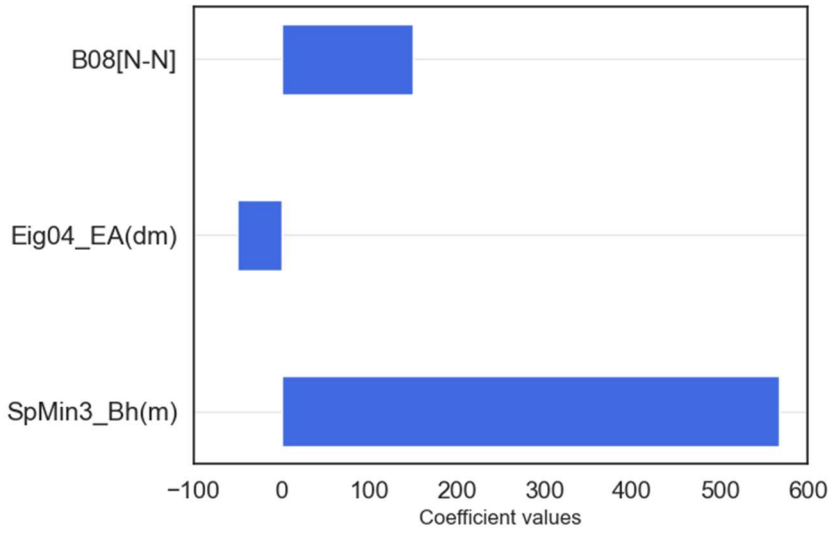

Figure 3. The magnitude of influence of different descriptors of 3-variable model on the the emission wavelength *(λ, nm)* according to **Eq. 7**.

The predictive ability of 3-variable model is $R^2 = 0.84$ for the training and $R^2 = 0.87$ for the test sets shows a very good correlation between predicted and observed $emission\ wavelength$ values for a set of 46 different iridium (III) complexes (**Fig. 4A**). Thus, it can be concluded that the 3-variable model is very robust.

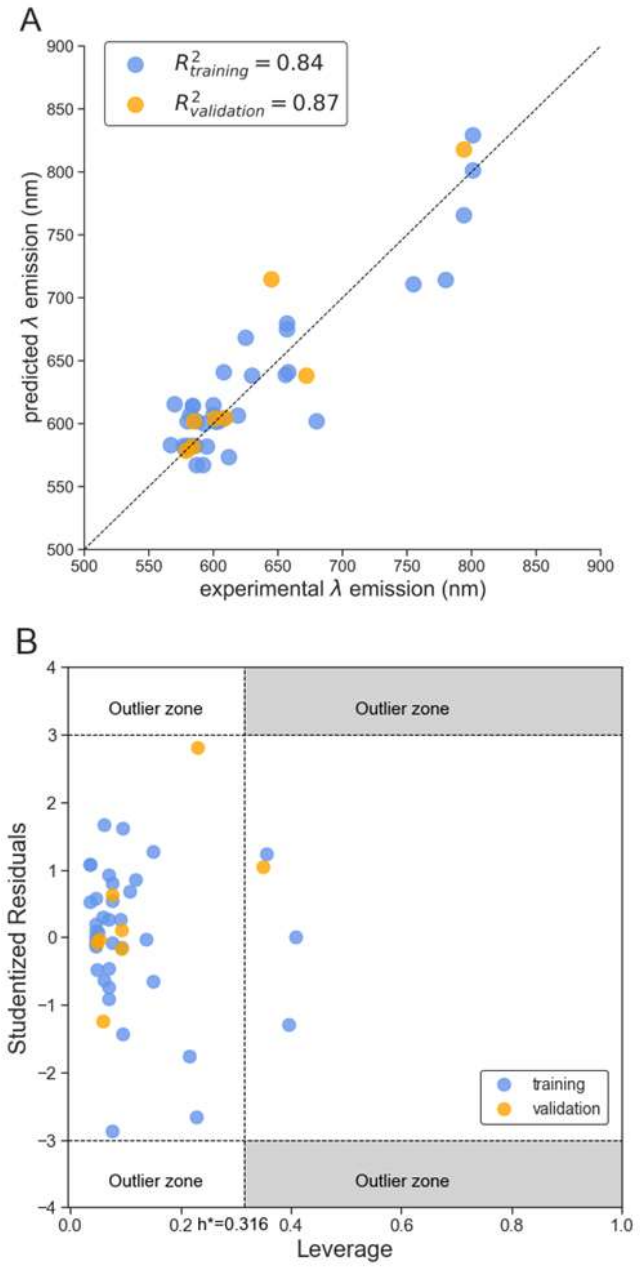

Figure 4. (A) The correlation plot between the observed and predicted values of $emission\ wave\ length\ (nm)$; **(B)** Williams plot of standardized residual versus leverage.

The Williams plot, presented in **Fig. 4B** for the 3-variable model, allows a graphical detection of both the outliers for the response and structurally different chemicals in the model. It consists of plotting the standardized residuals on the Y-axis and the leverage value on the X-axis.

Observation with standardized residuals greater than the (-3σ; +3σ) range is considered an outlier response. Leverage value represents the degree of influence that the structure of every single chemical has on the model. Williams plot shown in **Fig. 4B** confirms that all points are located at 3σ of error limit and reconfirms the absence of outliers, and all complexes considered for prediction in the 3-variable model fall within the applicability limit.

For further validation of the 3-variable model, we show the y-scrambling plot in **Fig. 5.** The y-scrambling plot is considered as a validation technique that underlines the robustness and uniqueness of the best model. The *Y*-scrambling plot is obtained by scrambling the *Y* values (experimental *emission wavelength*) randomly. In this randomization process, 2000 simulations per model are conducted. None of the random models (unrealistic models) show any good correlation in comparison with the selected 3-variable model: The selected model has significantly higher R^2 and Q^2 values than all other simulated random models (**Fig. 5**). Thus, the y-scrambling plot confirms that the developed model is a case of robustness and high correlation, rather than a mere coincidence. The essence of the above findings is that three detected physico-chemical and electro-topological parameters of the complexes (**Table 2**) play a pivotal role in determining the emission energy of Ir(III) complexes.

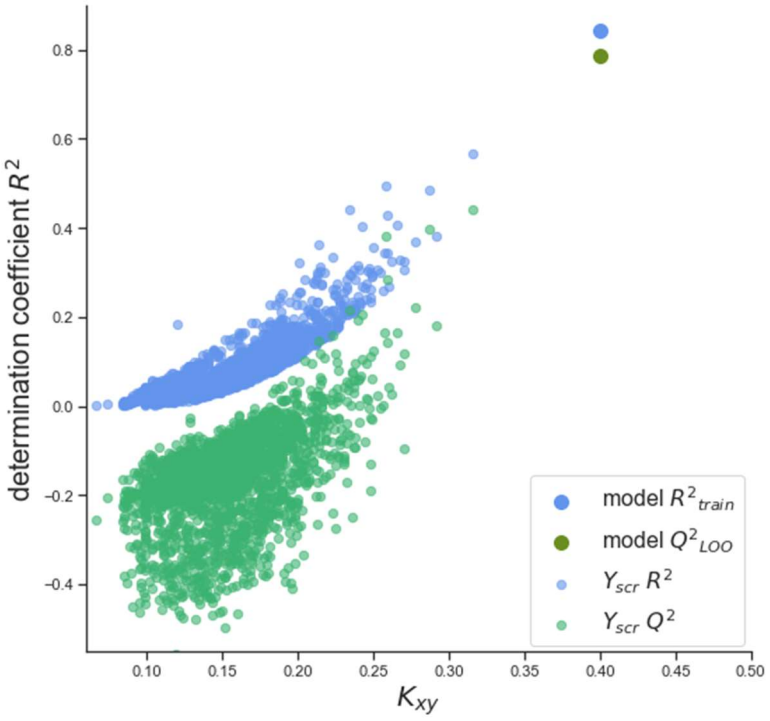

Figure 5. *Y*-scrambling plot, blue and green represent R^2 and Q^2 values of the original 3-value model (large circles) and random models (small circles).

One of the topological parameters from our model affecting the optical performance of Ir(III) complexes is the distance between nitrogens coordinated to the metal center (N-N topological distance in **Table 2**). To better understand this correlation, we closely consider eleven cationic Ir(III) complexes bearing similar N^N ligands coordinated to the metal center but with varyed electron-donating or electron-withdrawing substituents (complexes 1-5) at the 4-position of the 2-phenyl ring or different aromatic substituents with varied degrees of π-conjugation (complexes 6-11) on oxazolo[4,5-*f*][1,10]phenanthroline ligand. Schematic structures of these complexes are shown in **Fig. 6**.

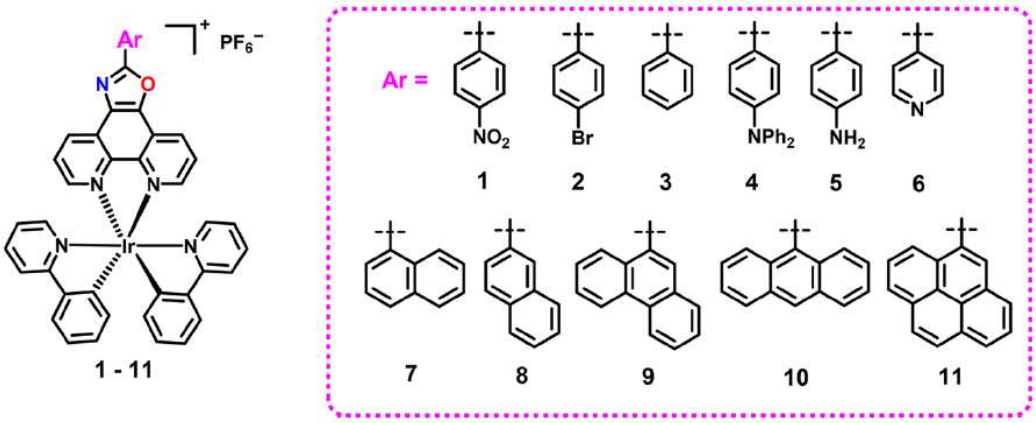

Figure 6. A set of structures for Ir(III) complexes 1-11 used to illustrate that the topological parameter, such as the N-N distance between coordinated nitrogens in oxazolo[4,5-f][1,10] phenanthroline ligand, directly corelates to the emission wavelength of these complexes.[13].

The N-N topological distance for these Ir(III) complexes as a function of their corresponding emission wavelength is shown in **Fig. 7**, demonstrating the inverse relationship between the N-N distance and the emission wavelength in all eleven complexes. The N-N topological distance of N^N ligands is changed due to both electron withdrawing/donating groups and a degree of π–conjugation of substituents. In particular, the N-N distance is reduced with increasing electron-withdrawing ability for complexes 1 and 2, and 6 (NO_2, Br, and pyridyl groups), as well as with decreasing of the conjugated length of substituents in complexes 6-11. A decrease in the N-N distance is accompanied by an increase in the emission wavelength (increased redshift). This correlation between the N-N distance of nitrogens coordinated with the metal center and the redshift in the emission energy can be rationalized by geometrical distortions in the octahedral ligand field contributed by substituents (e.g., shortening of the coordinated bonds laying at the same surface). This shortening of coordinated bonds results in stronger stabilization of the hole

state contributing to the lowest MLCT[3] transitions, compared to less distorted cases of complexes with electron donating ability (3-5) and larger π–conjugation of substituents (9-11). Thus, the complexes with smaller N-N distance show stronger redshifts in triplet emission.

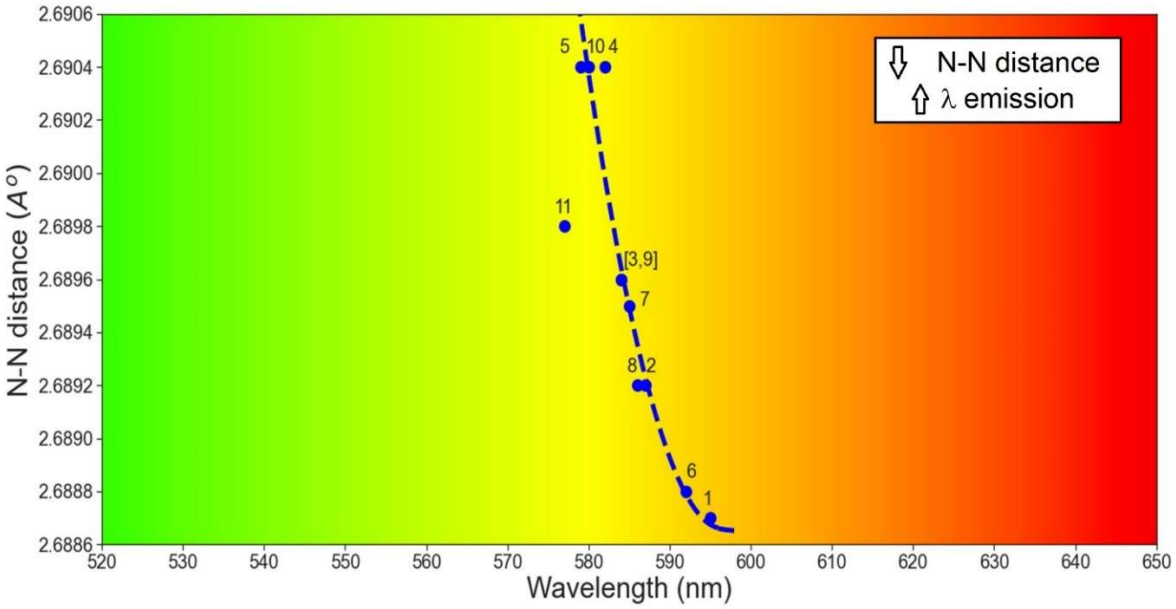

Figure 7. A relationship between the N-N topological distance and the wavelength of emission for the selected complexes.[13].

4. CONCLUSION

In conclusion, based on a data set of Ir(III) complexes the present study combines experimental testing, computational modeling and machine learning methodologies to study the optical properties of investigated complexes. We have successfully developed an interpretative machine learning-based QSAR model that reliably predicts emission wavelength and provides the foundations for theoretical evaluation of the optical properties of untested organometallic complexes, particularly Ir(III) complexes. The efficacy of the developed model was demonstrated

by high R^2 values for the training and test sets of 0.84 and 0.87, respectively, as well as by performing the validation using y-scrambling techniques.

Based on the developed ML/QSAR model we have formulated a hypothesis that mechanistically explains differences in emission wavelengths between structurally different individual Ir(III) complexes. To our knowledge, this is the first attempt to develop a predictive ML/QSAR model for optical properties of Ir(III) complexes. It was found, that there is a notable relationship between N^N distance in ligands of Ir(III) complexes and emission wavelengths. In this way, the combined experimental and computational approach showed great results in the current study and paves the way for the effective rational design of new organometallic complexes with desired optical properties.

Acknowledgments

Authors acknowledge the support from the National Science Foundation under grant number NSF CHE-1800476. This work is also supported in part by the NSF MRI award OAC-2019077, as well as ND EPSCoR award #IIA-1355466 and by the State of North Dakota. The authors thank Prof. Paola Gramatica for generously providing a free license for the QSARINS software. Authors also thank the Extreme Science and Engineering Discovery Environment (XSEDE) for the award allocation (TG-DMR110088). Supercomputing support from CCAST HPC System at NDSU is acknowledged.

Author contributions

A.K., G.M.C-M. and B.R. are designed and implemented the methodology. W.S. carried out empirical testing of the NIR emission properties of the organometallic complexes. L.L. designed and optimized molecular clusters for calculations. L.L., D.K. and S.K. performed quantum-mechanical calculations and analyzed the computational results. A.K., G.M.C-M. and B.R. selected the optimal structural descriptors, developed and validated the QSAR model, prepared the draft version of the manuscript and discussed the ML results. S.K, W.S., D.K. and B.R. discussed the overall results and edited the manuscript.

Additional information

The authors declare no competing financial interests. Supplementary information accompanies this paper at pubs.acs.com/jacs. Reprints and permission information is available online at http:// . Correspondence and requests for materials should be addressed to B.R.

REFERENCES

(1) Radwan, Y. K.; Maity, A.; Teets, T. S. Manipulating the Excited States of Cyclometalated Iridium Complexes with β-Ketoiminate and β-Diketiminate Ligands. *Inorganic Chemistry* **2015**, *54* (14), 7122-7131. DOI: 10.1021/acs.inorgchem.5b01401.

(2) Goswami, S.; Sengupta, D.; Paul, N. D.; Mondal, T. K.; Goswami, S. Redox Non-Innocence of Coordinated 2-(Arylazo) Pyridines in Iridium Complexes: Characterization of Redox Series and an Insight into Voltage-Induced Current Characteristics. *Chemistry - A European Journal* **2014**, *20* (20), 6103-6111. DOI: 10.1002/chem.201304369.

(3) Gärtner, F.; Cozzula, D.; Losse, S.; Boddien, A.; Anilkumar, G.; Junge, H.; Schulz, T.; Marquet, N.; Spannenberg, A.; Gladiali, S.; et al. Synthesis, Characterisation and Application of Iridium(III) Photosensitisers for Catalytic Water Reduction. *Chemistry - A European Journal* **2011**, *17* (25), 6998-7006. DOI: 10.1002/chem.201100235.

(4) Fan, C.; Zhu, L.; Jiang, B.; Li, Y.; Zhao, F.; Ma, D.; Qin, J.; Yang, C. High Power Efficiency Yellow Phosphorescent OLEDs by Using New Iridium Complexes with Halogen-Substituted 2-Phenylbenzo[d]thiazole Ligands. *The Journal of Physical Chemistry C* **2013**, *117* (37), 19134-19141. DOI: 10.1021/jp406220c.

(5) Tsuboyama, A.; Iwawaki, H.; Furugori, M.; Mukaide, T.; Kamatani, J.; Igawa, S.; Moriyama, T.; Miura, S.; Takiguchi, T.; Okada, S.; et al. Homoleptic Cyclometalated Iridium Complexes with Highly Efficient Red Phosphorescence and Application to Organic Light-Emitting Diode. *Journal of the American Chemical Society* **2003**, *125* (42), 12971-12979. DOI: 10.1021/ja034732d.

(6) Bolink, H. J.; Cappelli, L.; Coronado, E.; Grätzel, M.; Ortí, E.; Costa, R. D.; Viruela, P. M.; Nazeeruddin, M. K. Stable Single-Layer Light-Emitting Electrochemical Cell Using 4,7-Diphenyl-1,10-phenanthroline-bis(2-phenylpyridine)iridium(III) Hexafluorophosphate. *Journal of the American Chemical Society* **2006**, *128* (46), 14786-14787. DOI: 10.1021/ja066416f.

(7) Sun, L.; Galan, A.; Ladouceur, S.; Slinker, J. D.; Zysman-Colman, E. High stability light-emitting electrochemical cells from cationic iridium complexes with bulky 5,5′ substituents. *Journal of Materials Chemistry* **2011**, *21* (44), 18083. DOI: 10.1039/c1jm12984h.

(8) Lo, K. K.-W.; Chung, C.-K.; Lee, T. K.-M.; Lui, L.-H.; Tsang, K. H.-K.; Zhu, N. New Luminescent Cyclometalated Iridium(III) Diimine Complexes as Biological Labeling Reagents. *Inorganic Chemistry* **2003**, *42* (21), 6886-6897. DOI: 10.1021/ic0346984.

(9) Ma, D.-L.; Lin, S.; Wang, W.; Yang, C.; Leung, C.-H. Luminescent chemosensors by using cyclometalated iridium(iii) complexes and their applications. *Chemical science* **2017**, *8* (2), 878-889. DOI: 10.1039/c6sc04175b PubMed.

(10) Sun, J.; Wu, W.; Guo, H.; Zhao, J. Visible-Light Harvesting with Cyclometalated Iridium(III) Complexes Having Long-Lived ³ IL Excited States and Their Application in Triplet–Triplet-Annihilation Based Upconversion. *European Journal of Inorganic Chemistry* **2011**, *2011* (21), 3165-3173. DOI: 10.1002/ejic.201100501.

(11) Liu, B.; Lystrom, L.; Kilina, S.; Sun, W. Effects of Varying the Benzannulation Site and π Conjugation of the Cyclometalating Ligand on the Photophysics and Reverse Saturable Absorption of Monocationic Iridium(III) Complexes. *Inorganic Chemistry* **2019**, *58* (1), 476-488. DOI: 10.1021/acs.inorgchem.8b02714.

(12) Dragonetti, C.; Righetto, S.; Roberto, D.; Ugo, R.; Valore, A.; Fantacci, S.; Sgamellotti, A.; De Angelis, F. Cyclometallated iridium(iii) complexes with substituted 1,10-phenanthrolines: a new class of highly active organometallic second order NLO-phores with excellent transparency with respect to second harmonic emission. *Chemical Communications* **2007**, (40), 4116. DOI: 10.1039/b708073e.

(13) Zhu, X.; Cui, P.; Kilina, S.; Sun, W. Multifunctional Cationic Iridium(III) Complexes Bearing 2-Aryloxazolo[4,5-f][1,10]phenanthroline (N^N) Ligand: Synthesis, Crystal Structure, Photophysics, Mechanochromic/Vapochromic Effects, and Reverse Saturable Absorption. *Inorganic Chemistry* **2017**, *56* (22), 13715-13731. DOI: 10.1021/acs.inorgchem.7b01472.

(14) Liu, B.; Lystrom, L.; Cameron, C. G.; Kilina, S.; McFarland, S. A.; Sun, W. Monocationic Iridium(III) Complexes with Far-Red Charge-Transfer Absorption and Near-IR Emission: Synthesis, Photophysics, and Reverse Saturable Absorption. *European Journal of Inorganic Chemistry* **2019**, *2019* (16), 2208-2215, https://doi.org/10.1002/ejic.201900156. DOI: https://doi.org/10.1002/ejic.201900156 (acccessed 2022/11/29).

(15) Acharyya, J. N.; Desai, N. R.; Gangineni, R. B.; Vijaya Prakash, G. Effect of Photonic Cavity Interactions on Femtosecond Multiphoton Optical Nonlinear Absorptions from Bi2O3-Based One-Dimensional Photonic Crystal. *ACS Photonics* **2022**, *9* (6), 2092-2100, Article. DOI: 10.1021/acsphotonics.2c00326 Scopus.

(16) Yogeswari, C.; Hijas, K. M.; Girisun, T. C. S.; Nagalakshmi, R. Intensity dependent nonlinear absorption switching behavior of electrospun meta-nitroaniline nanofiber. *Optical Materials* **2020**, *100*, Article. DOI: 10.1016/j.optmat.2020.109691 Scopus.

(17) Zhu, X.; Lystrom, L.; Kilina, S.; Sun, W. Tuning the Photophysics and Reverse Saturable Absorption of Heteroleptic Cationic Iridium(III) Complexes via Substituents on the 6,6′-Bis(fluoren-2-yl)-2,2′-

biquinoline Ligand. *Inorganic Chemistry* **2016**, *55* (22), 11908-11919. DOI: 10.1021/acs.inorgchem.6b02028.
(18) Valore, A.; Cariati, E.; Dragonetti, C.; Righetto, S.; Roberto, D.; Ugo, R.; De Angelis, F.; Fantacci, S.; Sgamellotti, A.; Macchioni, A.; et al. Cyclometalated IrIII Complexes with Substituted 1,10-Phenanthrolines: A New Class of Efficient Cationic Organometallic Second-Order NLO Chromophores. *Chemistry - A European Journal* **2010**, *16* (16), 4814-4825. DOI: 10.1002/chem.200902788.
(19) Langhals, H. Color Chemistry. Synthesis, Properties and Applications of Organic Dyes and Pigments. 3rd revised edition. By Heinrich Zollinger. *Angewandte Chemie International Edition* **2004**, *43* (40), 5291-5292. DOI: 10.1002/anie.200385122.
(20) Mondal, S.; Seth, S. K.; Gupta, P.; Purkayastha, P. Ultrafast Photoinduced Electron Transfer between Carbon Nanoparticles and Cyclometalated Rhodium and Iridium Complexes. *The Journal of Physical Chemistry C* **2015**, *119* (44), 25122-25128. DOI: 10.1021/acs.jpcc.5b08633.
(21) Jin, C.; Liu, J.; Chen, Y.; Zeng, L.; Guan, R.; Ouyang, C.; Ji, L.; Chao, H. Cyclometalated Iridium(III) Complexes as Two-Photon Phosphorescent Probes for Specific Mitochondrial Dynamics Tracking in Living Cells. *Chemistry - A European Journal* **2015**, *21* (34), 12000-12010. DOI: 10.1002/chem.201501882.
(22) Castor, K. J.; Metera, K. L.; Tefashe, U. M.; Serpell, C. J.; Mauzeroll, J.; Sleiman, H. F. Cyclometalated Iridium(III) Imidazole Phenanthroline Complexes as Luminescent and Electrochemiluminescent G-Quadruplex DNA Binders. *Inorganic Chemistry* **2015**, *54* (14), 6958-6967. DOI: 10.1021/acs.inorgchem.5b00921.
(23) Liu, B.; Lystrom, L.; Kilina, S.; Sun, W. Tuning the Ground State and Excited State Properties of Monocationic Iridium(III) Complexes by Varying the Site of Benzannulation on Diimine Ligand. *Inorganic Chemistry* **2017**, *56* (9), 5361-5370. DOI: 10.1021/acs.inorgchem.7b00467.
(24) Wang, C.; Lystrom, L.; Yin, H.; Hetu, M.; Kilina, S.; McFarland, S. A.; Sun, W. Increasing the triplet lifetime and extending the ground-state absorption of biscyclometalated Ir(<scp>iii</scp>) complexes for reverse saturable absorption and photodynamic therapy applications. *Dalton Transactions* **2016**, *45* (41), 16366-16378. DOI: 10.1039/c6dt02416e.
(25) Sun, W.; Pei, C.; Lu, T.; Cui, P.; Li, Z.; McCleese, C.; Fang, Y.; Kilina, S.; Song, Y.; Burda, C. Reverse saturable absorbing cationic iridium(<scp>iii</scp>) complexes bearing the 2-(2-quinolinyl)quinoxaline ligand: effects of different cyclometalating ligands on linear and nonlinear absorption. *Journal of Materials Chemistry C* **2016**, *4* (22), 5059-5072. DOI: 10.1039/c6tc00710d.
(26) Wang, L.; Cui, P.; Kilina, S.; Sun, W. Toward Broadband Reverse Saturable Absorption: Investigating the Impact of Cyclometalating Ligand π-Conjugation on the Photophysics and Reverse Saturable Absorption of Cationic Heteroleptic Iridium Complexes. *The Journal of Physical Chemistry C* **2017**, *121* (10), 5719-5730. DOI: 10.1021/acs.jpcc.6b12947.
(27) Li, Z.; Cui, P.; Wang, C.; Kilina, S.; Sun, W. Nonlinear Absorbing Cationic Bipyridyl Iridium(III) Complexes Bearing Cyclometalating Ligands with Different Degrees of π-Conjugation: Synthesis, Photophysics, and Reverse Saturable Absorption. *The Journal of Physical Chemistry C* **2014**, *118* (49), 28764-28775. DOI: 10.1021/jp5073457.
(28) Pei, C.; Cui, P.; McCleese, C.; Kilina, S.; Burda, C.; Sun, W. Heteroleptic cationic iridium(<scp>iii</scp>) complexes bearing naphthalimidyl substituents: synthesis, photophysics and reverse saturable absorption. *Dalton Transactions* **2015**, *44* (5), 2176-2190. DOI: 10.1039/c4dt02384f.
(29) Lu, T.; Wang, C.; Lystrom, L.; Pei, C.; Kilina, S.; Sun, W. Effects of extending the π-conjugation of the acetylide ligand on the photophysics and reverse saturable absorption of Pt(<scp>ii</scp>) bipyridine bisacetylide complexes. *Physical Chemistry Chemical Physics* **2016**, *18* (41), 28674-28687. DOI: 10.1039/c6cp02628a.
(30) Wang, L.; Yin, H.; Cui, P.; Hetu, M.; Wang, C.; Monro, S.; Schaller, R. D.; Cameron, C. G.; Liu, B.; Kilina, S.; et al. Near-infrared-emitting heteroleptic cationic iridium complexes derived from 2,3-

diphenylbenzo[g]quinoxaline as in vitro theranostic photodynamic therapy agents. *Dalton Transactions* **2017**, *46* (25), 8091-8103. DOI: 10.1039/c7dt00913e.

(31) Pritchett, T. M.; Ferry, M. J.; Mott, A. G.; Shensky, W.; Haley, J. E.; Liu, R.; Sun, W. Long-lifetime reverse saturable absorption in a bipyridyl platinum(II) complex bearing naphthalimidylethynyl-substituted fluorenylacetylide ligands. *Optical Materials* **2015**, *39*, 195-198. DOI: 10.1016/j.optmat.2014.11.025.

(32) Li, Y.; Dandu, N.; Liu, R.; Kilina, S.; Sun, W. Synthesis and photophysics of reverse saturable absorbing heteroleptic iridium(<scp>iii</scp>) complexes bearing 2-(7-R-fluoren-2'-yl)pyridine ligands. *Dalton Trans.* **2014**, *43* (4), 1724-1735. DOI: 10.1039/c3dt52184b.

(33) Liu, R.; Dandu, N.; Chen, J.; Li, Y.; Li, Z.; Liu, S.; Wang, C.; Kilina, S.; Kohler, B.; Sun, W. Influence of Different Diimine (N^ΛN) Ligands on the Photophysics and Reverse Saturable Absorption of Heteroleptic Cationic Iridium(III) Complexes Bearing Cyclometalating 2-{3-[7-(Benzothiazol-2-yl)fluoren-2-yl]phenyl}pyridine (C^ΛN) Ligands. *The Journal of Physical Chemistry C* **2014**, *118* (40), 23233-23246. DOI: 10.1021/jp506765k.

(34) Li, Y.; Dandu, N.; Liu, R.; Li, Z.; Kilina, S.; Sun, W. Effects of Extended π-Conjugation in Phenanthroline (N^ΛN) and Phenylpyridine (C^ΛN) Ligands on the Photophysics and Reverse Saturable Absorption of Cationic Heteroleptic Iridium(III) Complexes. *The Journal of Physical Chemistry C* **2014**, *118* (12), 6372-6384. DOI: 10.1021/jp411259f.

(35) Liu, B.; Monro, S.; Lystrom, L.; Cameron, C. G.; Colón, K.; Yin, H.; Kilina, S.; McFarland, S. A.; Sun, W. Photophysical and Photobiological Properties of Dinuclear Iridium(III) Bis-tridentate Complexes. *Inorganic chemistry* **2018**, *57* (16), 9859-9872. DOI: 10.1021/acs.inorgchem.8b00789 PubMed.

(36) Li, Y.; Dandu, N.; Liu, R.; Hu, L.; Kilina, S.; Sun, W. Nonlinear Absorbing Cationic Iridium(III) Complexes Bearing Benzothiazolylfluorene Motif on the Bipyridine (NΛN) Ligand: Synthesis, Photophysics and Reverse Saturable Absorption. *ACS Applied Materials & Interfaces* **2013**, *5* (14), 6556-6570. DOI: 10.1021/am401133p.

(37) Jabeen, F.; Chen, M.; Rasulev, B.; Ossowski, M.; Boudjouk, P. Refractive indices of diverse data set of polymers: A computational QSPR based study. *Computational Materials Science* **2017**, *137*, 215-224. DOI: 10.1016/j.commatsci.2017.05.022.

(38) Karuth, A.; Alesadi, A.; Vashisth, A.; Xia, W.; Rasulev, B. Reactive Molecular Dynamics Study of Hygrothermal Degradation of Crosslinked Epoxy Polymers. *ACS Applied Polymer Materials* **2022**, *4* (6), 4411-4423. DOI: 10.1021/acsapm.2c00383.

(39) Varghese P J, G.; David, D. A.; Karuth, A.; Manamkeri Jafferali, J. F.; P M, S. B.; George, J. J.; Rasulev, B.; Raghavan, P. Experimental and Simulation Studies on Nonwoven Polypropylene-Nitrile Rubber Blend: Recycling of Medical Face Masks to an Engineering Product. *ACS omega* **2022**, *7* (6), 4791-4803. DOI: 10.1021/acsomega.1c04913 PubMed.

(40) Toropov, A. A.; Toropova, A. P.; Benfenati, E. QSAR-modeling of toxicity of organometallic compounds by means of the balance of correlations for InChI-based optimal descriptors. *Molecular Diversity* **2009**, *14* (1), 183-192. DOI: 10.1007/s11030-009-9156-6.

(41) Dems, M. A. E.; Laib, S.; Latelli, N.; Ouddai, N. A DFT-based Quantitative structure activity relationship Study of organometallic estradiol derivatives. *Journal of Chemical and Pharmaceutical Sciences* **2017**, *10*.

(42) Chen, M.; Jabeen, F.; Rasulev, B.; Ossowski, M.; Boudjouk, P. A computational structure–property relationship study of glass transition temperatures for a diverse set of polymers. *Journal of Polymer Science Part B: Polymer Physics* **2018**, *56* (11), 877-885. DOI: 10.1002/polb.24602.

(43) Ahmed, L.; Rasulev, B.; Turabekova, M.; Leszczynska, D.; Leszczynski, J. Receptor- and ligand-based study of fullerene analogues: comprehensive computational approach including quantum-chemical, QSAR and molecular docking simulations. *Organic & Biomolecular Chemistry* **2013**, *11* (35), 5798. DOI: 10.1039/c3ob40878g.

(44) Christenholz, C. L.; Obenchain, D. A.; Peebles, R. A.; Peebles, S. A. Rotational Spectroscopic Studies of C–H···F Interactions in the Vinyl Fluoride···Difluoromethane Complex. *The Journal of Physical Chemistry A* **2014**, *118* (9), 1610-1616. DOI: 10.1021/jp500312r.
(45) Ernzerhof, M.; Scuseria, G. E. Assessment of the Perdew–Burke–Ernzerhof exchange-correlation functional. *The Journal of Chemical Physics* **1999**, *110* (11), 5029-5036. DOI: 10.1063/1.478401.
(46) Perdew, J. P.; Yang, W.; Burke, K.; Yang, Z.; Gross, E. K. U.; Scheffler, M.; Scuseria, G. E.; Henderson, T. M.; Zhang, I. Y.; Ruzsinszky, A.; et al. Understanding band gaps of solids in generalized Kohn-Sham theory. *Proceedings of the National Academy of Sciences of the United States of America* **2017**, *114* (11), 2801-2806. DOI: 10.1073/pnas.1621352114 PubMed.
(47) Wadt, W. R.; Hay, P. J. Ab initio effective core potentials for molecular calculations. Potentials for main group elements Na to Bi. *The Journal of Chemical Physics* **1985**, *82* (1), 284-298. DOI: 10.1063/1.448800.
(48) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *The Journal of Chemical Physics* **1982**, *77* (7), 3654-3665. DOI: 10.1063/1.444267.
(49) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *The Journal of Chemical Physics* **1980**, *72* (1), 650-654. DOI: 10.1063/1.438955.
(50) Barone, V.; Cossi, M.; Tomasi, J. Geometry optimization of molecular structures in solution by the polarizable continuum model. *Journal of Computational Chemistry* **1998**, *19* (4), 404-417. DOI: 10.1002/(sici)1096-987x(199803)19:4<404::aid-jcc3>3.0.co;2-w.
(51) *ChemSketch, A. C. D.*; Advanced Chemistry Development. Inc: Toronto, Canada, 2015. (accessed.
(52) Toddeschini; R.; Consonni, V.; Mauri, A.; Pavan, M. Dragon Software for the Calculation of Molecular Descriptors, Version 6 for Windows; Talete SRL: Milan, Italy. 2014.
(53) Todeschini, R.; Consonni, V. *Molecular Descriptors for Chemoinformatics*; Wiley, 2009. DOI: 10.1002/9783527628766.
(54) Devillers, J. Genetic Algorithms in Computer-Aided Molecular Design. In *Genetic Algorithms in Molecular Modeling*, Elsevier: 1996; pp 1-34.
(55) Gramatica, P.; Cassani, S.; Chirico, N. QSARINS-chem: Insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS. *Journal of Computational Chemistry* **2014**, *35* (13), 1036-1044. DOI: 10.1002/jcc.23576.
(56) Karuth, A.; Alesadi, A.; Xia, W.; Rasulev, B. Predicting glass transition of amorphous polymers by application of cheminformatics and molecular dynamics simulations. *Polymer* **2021**, *218*, 123495. DOI: https://doi.org/10.1016/j.polymer.2021.123495.
(57) Gramatica, P.; Giani, E.; Papa, E. Statistical external validation and consensus modeling: A QSPR case study for Koc prediction. *Journal of Molecular Graphics and Modelling* **2007**, *25* (6), 755-766. DOI: 10.1016/j.jmgm.2006.06.005.