Data-driven Strategies to Predict Ruthenium Complex Photocleavage Efficiency

Jinge Xu

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2021

Committee:

Cole DeForest

David Beck

Program Authorized to Offer Degree:

Chemical Engineering

©Copyright 2021
Jinge Xu

University of Washington

Abstract

Data-driven Strategies to Predict Ruthenium Complex Photocleavage Efficiency

Jinge Xu

Chair of the Supervisory Committee:
Cole DeForest
Department of Chemical Engineering

Ruthenium (Ru) complex-based photocages have shown great potential in photochemotherapy due to their activation by red- and infrared light. However, different Ru complexes exhibit dramatically varying quantum yields in the photocleavage process. To optimize the structures for a high quantum yield, in this work, we developed machine learning models and density-functional theory (DFT) protocols to predict the quantum yield for photocleavage of Ruthenium complexes. The built machine learning models were not accurate enough to classify the quantum yield for a given Ru complex structure. An alternative DFT protocol was designed to bridge the experimental quantum photorelease quantum yield and computational energy gap between the 3MLCT and 3MC states. This approach demonstrated the energy gap between the 3MLCT state and the 3MC state can serve as a ‘barrier’ in the ligand dissociation process and could help predict the quantum yield of photolabile Ru-based complexes.

Data-driven Strategies to Predict Ruthenium Complex Photocleavage Efficiency

Abstract

Ruthenium (Ru) complex-based photocages have shown great potential in photochemotherapy due to their activation by red- and infrared light. However, different Ru complexes exhibit dramatically varying quantum yields in the photocleavage process. To optimize the structures for a high quantum yield, in this work, we developed machine learning models and density-functional theory (DFT) protocols to predict the quantum yield for photocleavage of Ruthenium complexes. The built machine learning models were not accurate enough to classify the quantum yield for a given Ru complex structure. An alternative DFT protocol was designed to bridge the experimental quantum photorelease quantum yield and computational energy gap between the ^{3}MLCT and 3MC states. This approach demonstrated the energy gap between the ^{3}MLCT state and the ^{3}MC state can serve as a 'barrier' in the ligand dissociation process and could help predict the quantum yield of photolabile Ru-based complexes.

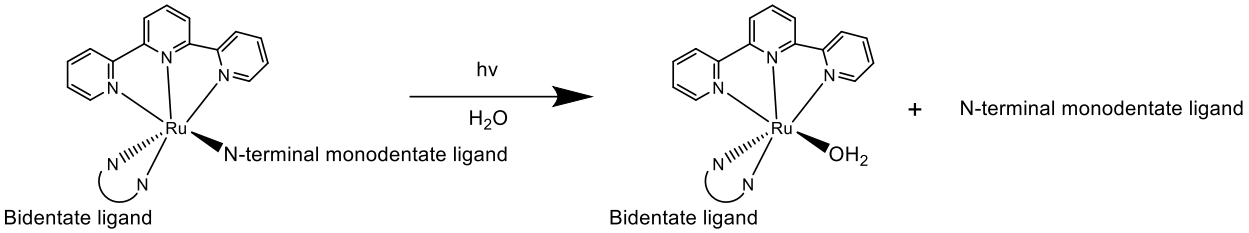
Introduction

Photolabile systems have been used in a wide range of research fields, including solar energy conversion materials[1], biomedical research[2–6], and drug delivery[7–14]. In particular, by incorporating functional biomolecules or small molecule drugs, photoactivated molecules can release bioactive moieties through photochemical transmission, usually termed 'photocage'. Photocaging is a highly attractive method that enables rapid, precise, and spatiotemporal control over biological activities under light irradiation[11,15], and is currently being explored as a new platform in on-demand therapeutics and diagnostics in localized area[5,16,17]. 2-nitrobenzyl[18,19], coumarin[20,21], benzoin[22], p-hydroxyphenacyl[23], boron dipyrromethene[11,14], and ruthenium complexes[24–26] have been modified and designed as photocages. Researchers have tried to install various bioactive molecules on photocages, including proteins[7,27], peptides[3], API[28], DNA[29,30], and RNA[31].

While organic photocages are traditionally used, they typically require UV light to trigger photoreaction that does not readily penetrate deeply into human body and cause severe damage to human skin, tissue, and even DNA[32,33]. In contrast, transition metal complexes, especially ruthenium complexes, can be activated by red light or near-infrared light due to their strong charge-transfer absorption in the visible light range[10,26]. In addition, a bond between transition metals and the ancillary ligands is usually weaker than an organic sigma bond[34]. Thereby, ruthenium complexes enable a facile release of functional compounds activated by red light or near-infrared light that penetrates deeper into tissues with less concern over potential DNA damage.

Generally, Ru complexes are composed of a ruthenium cation, two ancillary ligands, and a monodentate caged ligand. Under irradiation, valid Ru photocages can be excited to the singlet metal-to-ligand charge transfer state (^{1}MLCT) from the ground state (S_0). Through a transient intersystem crossing, they transform to the lowest-lying triplet metal-to-ligand charge transfer state

(^{3}MLCT), with another stable excited state (^{3}MC) close enough to be thermally populated[35]. The triplet metal-centered state (^{3}MC or ^{3}LF) plays an essential role in transition metal-based photochemical devices or photochemotherapy due to their dissociative property[36]. Once the Ru complexes convert to the dissociative ^{3}MC state, the caged bioactive molecule would be substituted by a solvent molecule, accomplishing the mission of drug delivery, as shown in **Scheme 1**.



Scheme 1 Light-triggered ligand dissociation of Ru complexes in H_2O

Although Ru complexes are showing great potential in photochemotherapy, several limitations remain that hinder their widespread clinical application. First, most published Ru complexes cannot meet the requirement for a high quantum efficiency of photo-induced ligand release[37]. Furthermore, several possess considerable cytotoxicity even in the intact form in the dark. Further complicating matters, many Ru photocages exhibit poor water stability. Finally, their thermal stability is not often evaluated systematically. Computational methods may shed new light on development of newly useful species.

In photochemistry, quantum yield (QY) can assess the quantum efficiency of photoreaction. In the photo-induced ligand dissociation process, quantum yield is defined as the ratio of the number of photons emitted to the number of photons absorbed. Since Ru(II) complexes usually undergo a single-photon activation, the quantum yield also equals the number of decomposed molecules to the number of excited molecules[38].

Besides in vitro and in vivo experiments, in silico methods also serve as powerful tools in revealing sophisticated chemical mechanisms and exploring novel structures. In the past decades, density functional theory (DFT) has emerged as the most popular electronic structure method in computational chemistry. DFT simulations can provide quantum chemical level insight to understand the compound behaviors and design new materials[39,40]. The core of DFT calculations is to solve the Schrödinger equation of a collection of atoms from first principles, which achieve high accuracy at the cost of substantial computational cost and appropriate choice of wavefunction and basis set. Owing to their outstanding performance in multiscale regression and classification with relatively cheap computational cost, machine learning techniques have been widely used over the last few years to predict the chemical properties and decode the key attributing structures[41–43].

Machine learning models are most extensively applied to develop quantitative structure-property relationship (QSPR) models[44–46]. QSPR models can reveal a mathematical relationship between structural attributes and property of interest at the quantum chemical level. They are used to predict targeted properties for a wide range of chemicals prior to or in lieu of costly and labor-intensive experimental measurements, which is enticing and promising in biological and pharmaceutical science and engineering. The first step in the QSPR model development process is to generate a valid database of data collected in similar environments. After the data generation, molecular descriptors are gathered as the features in the machine learning model. Molecular fingerprints, physicochemical descriptors, and molecular graphs are the most widely used descriptors. Feature

selection is a critical part of a superior model. The features are supposed to embody the difference of structures while keeping a limited size. Then the machine learning algorithms are applied for dataset training, validation, and testing, where parameters need to be tuned to optimize models' performance. Error functions are criteria to assess the model's adaption goodness. Before a qualified model is built, several model adjustments and feature updates may be required to better fit the dataset[44].

Given that the expensive and time-consuming workflow of synthesizing the Ru(II) complexes and measuring their photo-reaction quantum yield, in this work, we developed machine learning models to predict the photo-induced ligand release quantum yield of given Ru(II) complexes. In addition, since most of the Ru(II) complexes are synthesized in the lab and not registered yet, there are insufficient molecular descriptors for Ru(II) complexes as an intact compound, in both quantity and variety. Therefore, DFT calculations were also performed to provide an alternative for the machine learning models.

Methodology

In this work, the data processing was performed with Python 3.7. Examples and visualizations were organized in the Jupyter Notebook (Anaconda 3), and customed packages were compiled in Visual Studio Code. DFT calculations were simulated on Gaussian 16 b.01 and visualized in Gauss View 6.1.

Data Generation

51 Ru(II) complexes' structural composition and photorelease quantum yield value (QY) were gathered from the published literature. Since the QY value is sensitive to the experimental condition, we confirmed the standard characterizing environment to be under 470 nm wavelength irradiation in water, which is the most common environment among the dataset. For the experiments conducted under a rare environment, the data were discarded. Because of the scarcity of the QY data and the uncertainty in the experiment process, we identified the task to be a binary classification. By setting the threshold at 0.03, a QY below the threshold was labeled as "low quantum yield", a QY above the threshold was labeled as "high quantum yield".

Feature Selection

Feature selection, referring to molecular representation in this work, is the crucial section in the model development. Considering the lack of chemical properties for Ru(II) complexes as an intact compound and the common three-arm coordination of the Ru(II) complexes, a Ru(II) complex was conceived as the combination of three parts: ligand 1 (tridentate ancillary ligand), ligand 2 (bidentate ancillary ligand) and ligand 3 (caged/released ligand). First, the ligands' structure was converted to their canonical simplified molecular-input line-entry system (SMILES), a specification in the form of a line notation for describing the structure of chemical species using short ASCII strings. Then the Molecular ACCess System (MACCS) keys of each SMILES string were calculated, and according to the fragments segregation, sequentially assembled into a new 498-bit binary string as the elementary molecular descriptors. Because of the overfull features (498) compared to the samples (51), all the features with a zero-variance was abandoned. 180 bits remained as the features in the machine learning model.

In this project, MACCS keys were chosen as the molecular descriptor. MACCS keys are 166-bit two-dimensional structure molecular fingerprints that encode substructures into binary bit string

to represent their presence or absence in the molecule of interest[47]. MACCS keys possess clearly defined documentation for every bit, facilitating the pattern retrieval.

Model Generation and Validation

In this binary classification task, several algorithms were applied, including artificial neural network(ANN), random forest(RF), and support vector machine(SVM) with linear kernel and rbf kernel. The model packages were imported from the scikit-learn library, and parameters were tuned in the training process. Model performance was evaluated in their predictive accuracy, training error, and testing error, taking the average of 100-runs of each algorithm. The dataset was randomly divided into a 70% size training set and the a 30% testing set in every run. A fully connected 100-node single hidden layer neural network was applied, with relu activation function and lbfgs solver. Both linear kernel and rbf kernel experimented on SVM. The predictive accuracy was defined as the ratio of the size of correctly classified samples over the size of the test dataset. Log loss function was employed as the error function to assess the goodness of fit. The log loss is defined as the negative log-likelihood of a logistic model that returned `y_pred` for its training data `y_true`. For a single sample with true label $y \in \{0,1\}$ and a probability estimate p=Pr(y=1), the log loss is:

$$L_{log}(y,p) = \left(-ylog(p) + (1-y)log(1-p)\right)$$

DFT Configuration

A workflow was designed to calculate the energy gap between the ^{3}MLCT state and the ^{3}MC state. Initially, a rational geometry guess for Ru(II) complexes was generated and optimized for the geometry in the ground state, following by another geometry optimization by modifying the multiplicity to obtain the geometry in the triplet state. Then a vertical absorption was performed to search for ^{3}MLCT state. An initial guess for geometry in the ^{3}MC state was generated by securely long elongating the bond between Ru atom and N atom in the caged ligand to imitate the photorelease process. In view of the reorganization of ^{3}MC state geometry, a geometry optimization was performed to obtain the geometry in the ^{3}MC state. Finally, a vertical absorption was calculated to search for the ^{3}MC state, and the energy gap between the ^{3}MLCT state and the ^{3}MC state was computed.

BP86 (Becke88 + Perdew86) was adopted as the wave function. It has been reported BP86 could provide smallest energy difference between the ^{3}MLCT state and the ^{3}MC state[48]. A mixed basis set was adopted. SDD basis set, along with SDD effective core potential, was used on Ru atoms for its good fitness in representing transition metal atoms. The usage of effective core potential (ECP) can reduce computational cost on multiple metal center systems. In our work, the simulated structures were all single Ru center systems, but I found the SDD ECP could facilitate the geometry optimization convergence; therefore we kept it. 6-31G(d) basis set was used for H, C, and N atoms and 6-31G(d,p) basis set was used for O atoms. 6-31G is a good basis set in representing light atoms. Although it is relatively crude compared to other advanced 3-zeta basis sets, it was adequate for our calculations and reduced computational cost. The usage of polarization function(d function for H, C, N, d and p function for O) can improve the precise representation for bonds. In the excited states(3MLCT, 3MC) searching process, an ideal vertical adsorption calculation was adopted to save tons of computational resources. It has been demonstrated the overestimation of the excitation energy from vertical absorption can be acceptable by carefully choosing wavefunction and basis set[49]. Time-dependent DFT calculation was performed, and an implicit SMD solvation model with

methanol solvent was used in excited states searching. All the geometry optimizations and TDDFT calculations were under 298.15K and 1atm.

Simulated Ru Complexes

Two Ru complexes systems, a 'tpy' system, and a 'dqpy' system, were investigated. Ru complexes were three-coordinated structures in both systems: a tridentate ancillary ligand, a bidentate ancillary ligand, and a monodentate caged ligand. For the Ru complexes in the 'tpy' system, the tridentate ligand was constantly 2,2';6',2''-terpyridine, with a series combination of bidentate ligand and caged ligand: 2,2′-bipyridine + pyridine(1), 6,6′-dimethyl-2,2′-bipyridine + pyridine(2), 2,2′-bipyridine + acetonitrile(3), 2,9- dimethyl-1,10-phenanthroline + pyridine(4), 2,2′-bipyridine + pyrazine(10), 2,9- dimethyl-1,10-phenanthroline + pyrazine(13), 2,2′-biquinoline + pyridine(14). The quantum yield data for 'tpy' system were determined under 470nm irradiation in H_2O. For the Ru complexes in the 'dqpy' system, the tridentate ligand and the caged ligand were fixed at 2,6-di(quinolin-2-yl)pyridine and acetonitrile, respectively. And the bidentate ligands included 2,2′-bipyridine(6), 6,6′-dimethyl-2,2′-bipyridine(7), 1,10-phenanthroline(8), and acetylacetonate(9). The quantum yield data for 'dqpy' system were determined under 450nm irradiation in 5% acetone solution. The structures are shown in **Figure 1**.

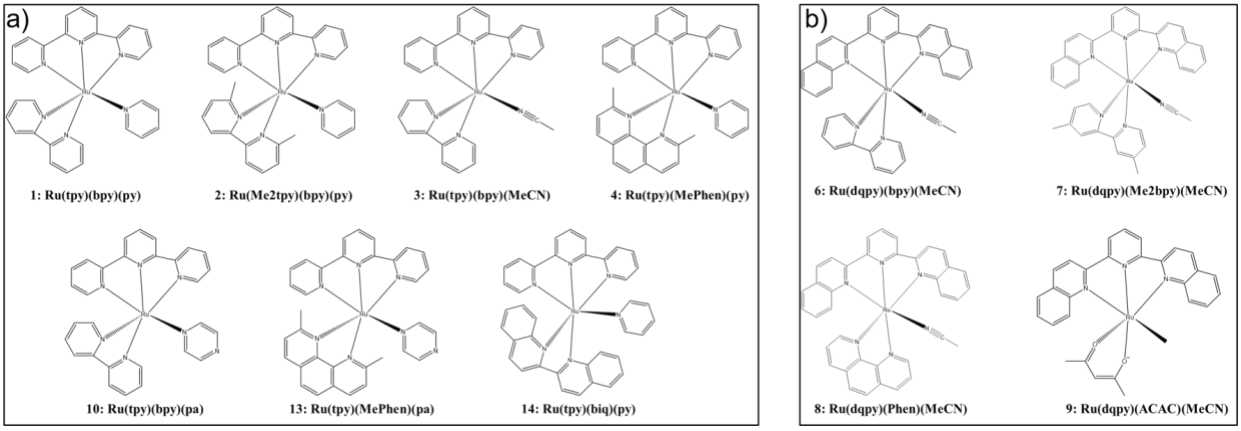


Figure 1 Simulated structures in a) 'tpy' system and b) 'dqpy' system

Results and Discussions

a)

	model	predictive accuracy	training error	testing error
0	ANN	0.635	1.524	12.595
1	RF	0.704	1.636	10.224
2	SVM(linear)	0.709	3.230	10.039
3	SVM(rbf)	0.779	4.917	7.645

b)

	model	predictive accuracy	training error	testing error
0	ANN	0.727	2.875	9.441
1	RF	0.725	2.926	9.487
2	SVM(linear)	0.745	3.982	8.796
3	SVM(rbf)	0.771	5.147	7.906

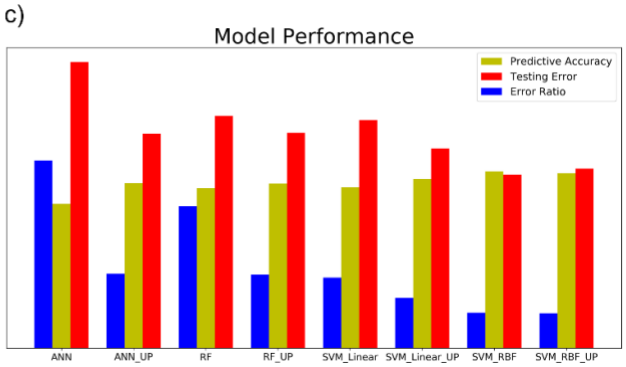


Figure 2 Machine learning models' predictive accuracy, training and testing error (a) trained with original features (b) trained with updated features (c)model performance comparison before and after feature update

Figure 2a listed the results when the reduced 180-bit MACCS keys were used as the feature. The SVM(rbf kernel) model exhibited the best predictive accuracy(0.779) and the lowest testing error(7.635) among the four models. In contrast, ANN had a low predictive accuracy(0.635) and a large testing error(12.595), suggesting ANN's poor performance on the dataset. Nevertheless, a severe overfitting problem was observed in all the models: their testing errors were much higher compared to the training error(8 fold for ANN, 6 fold for RF, 3 fold for SVM(linear kernel), and 1.5 fold for SVM(rbf kernel)). To mitigate the overfitting problem, a feature update was performed. During the training process of both RF and SVM(linear kernel), the feature importance was reported in every run. The feature was labeled "important feature" when the individual importance was above the threshold at 0.1. After the training process, the features labeled "important feature" in over 50% runs were gathered. Features gathered from both RF and SVM(linear kernel) were selected as the update features. As a result, 20 overlapped MACCS keys bits were selected, and the four models were retrained with the update feature. **Figure 2b** shows the retrained models' performance, and **Figure 2c** compares the predictive accuracy, testing error, and testing/training error ratio of models before and after the feature update. After the feature update, there were apparent elevations in predictive accuracy and reduction in testing error of ANN, RF, and SVM(linear kernel), owing to the abandonment of trivial features.

In contrast, SVM(rbf kernel) exhibited negligible variation after the feature update and still possessed the highest predictive accuracy and lowest testing error. Such stability can attribute to the fitting mechanism of SVM(rbf kernel): it finds the best non-linear hyperplane by maximizing the margin of data from both labels. **Figure 3** displayed the data distribution in Principal Component 1(PC1) and Principal Component 2(PC2) plane after performing principal component analysis.

Principal component analysis (PCA) projects a high-dimensional feature onto a visible two-dimensional subspace, where the variance of the dataset is maximized. By visualizing the relative locations of the samples in the chosen principal component subspace, we can obtain a general sense of the data distribution in the feature space. As shown by **Figure 3b**, the majority of 'high quantum yield' and 'low quantum yield' were basically separated, accounting for the better performance and stability of SVM (rbf kernel) in this binary classification task. Meanwhile, Figure 3 also suggested some outliers and overlaps in the PC1-PC2 subspace, which could cause considerable impact on the training and testing process, limiting further model improvement. These undesired distributed data might result from the roughness of feature (MACCS keys). MACCS keys were chosen as features due to their well-defined documentation and shorter bitstring. However, the shorter bitstring might sacrifice some specific detail in distinguishing the minor structural difference. Thus, MACCS keys might not be the proper molecular descriptors in this dataset. In addition, the small size of the dataset is another factor limiting the model performance. The train test split would lead to a smaller set, increasing the uncertainty and fortuity in the training process. Therefore, the enrichment of the dataset and enhancement of feature selection were sought to further develop a qualified model.

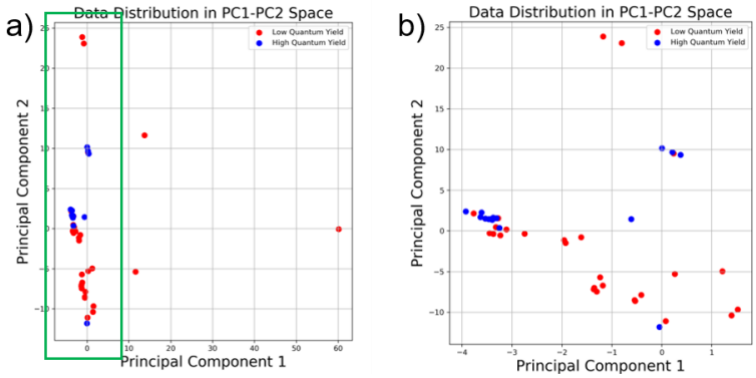


Figure 3 Data distribution in principal component 1 and principal component 2 subspace a) in the whole subspace b) in the localized area(green box marked in a))

Due to the unsatisfying machine learning model development, an alternative was pursued to predict the photorelease quantum yield based on Ru(II) complexes' structures. Given the transformation of the excited state during the photorelease process, we hypothesized that the energy gap between ^{3}MLCT state and ^{3}MC state served as the barrier for photo-induced ligand release: a larger energy gap would increase the difficulty for the photorelease, while a smaller energy gap could facilitate the ligand release, increasing the quantum yield in the photorelease process.

Table 1 Simulated structures[a] and their energy in ^{3}MLCT state and ^{3}MC state, energy gap, reported experimental photorelease quantum yield and their belongings

No	Structure	3MLCT energy/eV	3MC energy/eV	Energy gap/eV	Reported QY	System
1	Ru(tpy)(bpy)(py)	-1580.994558	-1580.964796	0.029762	0.0000	tpy' system
2	Ru(tpy)(Me2bpy)(py)	-1659.608612	-1659.603174	0.005438	0.1040	tpy' system
3	Ru(tpy)(bpy)(MeCN)	-1465.459840	-1465.434337	0.025503	0.0125	tpy' system
4	Ru(tpy)(MePhen)(py)	-1735.842095	-1735.830449	0.011646	0.0580	tpy' system
6	Ru(dqpy)(bpy)(MeCN)	-1772.727171	-1772.714263	0.012908	0.0039	dqpy' system
7	Ru(dqpy)(Me2bpy)(MeCN)	-1851.363417	-1851.348887	0.014530	0.0032	dqpy' system
8	Ru(dqpy)(Phen)(MeCN)	-1848.960257	-1848.947361	0.012896	0.0048	dqpy' system
9	Ru(dqpy)(ACAC)(MeCN)	-1622.694378	-1622.694101	0.000277	0.0096	dqpy' system
10	Ru(tpy)(bpy)(pa)	-1597.024398	-1596.998472	0.025926	0.0013	tpy' system
13	Ru(tpy)(MePhen)(pa)	-1751.877027	-1751.871728	0.005299	0.1320	tpy' system
14	Ru(tpy)(biq)(py)	-1888.264317	-1888.248886	0.015431	0.0140	tpy' system

[a]The chemical structures inside the parentheses refered to the ligands. `tpy`=2,2';6',2"-terpyridine, `bpy`=2,2′-bipyridine, `py`=pyridine, `Me2bpy`=6,6′-dimethyl-2,2′-bipyridine, `MeCN`=acetonitrile, `MePhen`=2,9- dimethyl-1,10-phenanthroline, `dqpy`=2,6-di(quinolin-2-yl)pyridine, `Phen`=1,10-phenanthroline, `ACAC`=acetylacetonate, `pa`=pyrazine, `biq`=2,2′-biquinoline

The structural information and energies for ^{3}MLCT and ^{3}MC states were listed in **Table 1**. For the Ru(II) complexes, optimized geometry in the ground state and ^{3}MLCT state was unique, while there were multiple possible geometries for the ^{3}MC state by stretching different Ru-N bonds. Nisbett et al. have already demonstrated the optimized excited states geometry obtained by stretching the bond between Ru and leaving group possessed the lowest energy, thereby being the most stable geometry and consistent with the experimental results monodentate ligand that dissociated with Ru under irradiation[48]. Therefore, in this work, we stretched the monodentate ligands far from the Ru center (4-6 fold original bond length) to obtain the initial guess and then

applied a geometry optimization for geometry in the ^{3}MC state. The calculated energy gaps and experimental photorelease quantum yields are displayed in **Figure 4**. Aligned with the hypothesis, there is a reverse relationship observed in both systems: the larger the energy gap calculated, the lower the quantum yield measured, indicating the energy gap between ^{3}MLCT and ^{3}MC state can be used as a proxy for photorelease quantum yield. Remarkably, the counter-response exhibited a smoother adaption in the high quantum yield region, illustrating promising reliability in predicting the high quantum yields, which is the main interest of the entire work. In the low quantum yield region, even though the hypothesis was validated, the correlation was not ideally linear or exponential, likely owing to the trade-off of the DFT configuration. To save computational resources, vertical absorption was adopted to search for the excited states. The adequate basis set fulfilled the requirement to calculate the energy gap while can still mislead in some cases that there was only a minor energy gap difference between two structures, which could account for the roughness and sharpness in the low quantum yield area. The `barrier` role of the energy gap in the photorelease process can be regarded as a quantification for the degree of difficulty of transforming from ^{3}MLCT and ^{3}MC state. Although it is acknowledged that the Ru(II) complexes would undergo the transformation of the excited state to release the ligand, there is no literature formulating the difficulty of the conversion yet. Here, we demonstrate a reverse relationship that bridges the experimental and computational work. The relationship can help predict the quantum yield with a given Ru(II) complex structure, currently valid for the energy gap higher than 0.00529 Hartree (quantum yield lower than 0.132).

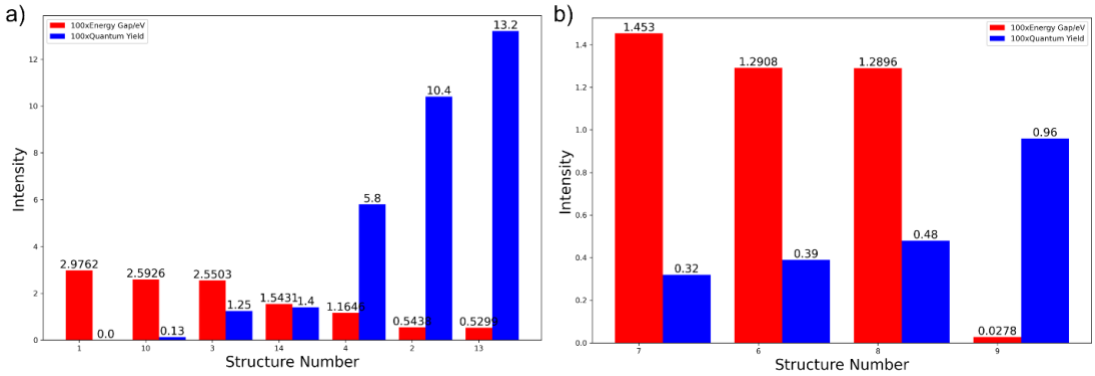


Figure 4 Bar plots of calculated energy gaps and experimental quantum yields for structures in (a)`tpy` system (b) `dqpy` system

Conclusions

In conclusion, in order to predict the photorelease quantum yield of Ru(II) complexes, several machine learning models, including ANN, RF, and SVM, were adopted, and MACCS keys were used as features. Models exhibited comparable performance and a common overfitting problem since MACCS keys might not be exact enough to enlarge the minor difference in the similar Ru(II) complexes structures. Therefore, a better molecular representation and dataset enrichment was being sought. Meanwhile, a surrogate DFT protocol was developed and estimated the experimental quantum yield by computing the energy gap between the lowest-lying ^{3}MLCT and the dissociative ^{3}MC state under specific circumstances. In particular, the DFT model worked better in a high quantum yield region with a smoother adaption, revealing the model's promising prediction application on the Ru complexes photocages.

Acknowledgement

The author gratefully acknowledges the supervision of Prof. Cole DeForest and Prof. David Beck from University of Washington. Thanks for Dr. Teresa Rapp's knowledge support on Ru complexes and Andrew Wildman's technical support for gaussian and hyak. The author also thanks the Research Computing Club at University of Washington for software and computing support.

References

1. Kim, J. Y., Lee, J. W., Jung, H. S., Shin, H. & Park, N. G. High-Efficiency Perovskite Solar Cells. *Chem. Rev.* **120**, 7867–7918 (2020).
2. Agarwal, H. K. *et al.* Calcium Uncaging with Visible Light. *J. Am. Chem. Soc.* **138**, 3687–3693 (2016).
3. DeForest, C. A. & Anseth, K. S. Photoreversible patterning of biomolecules within click-based hydrogels. *Angew. Chemie - Int. Ed.* **51**, 1816–1819 (2012).
4. Deforest, C. A. & Anseth, K. S. Advances in bioactive hydrogels to probe and direct cell fate. *Annu. Rev. Chem. Biomol. Eng.* **3**, 421–444 (2012).
5. Badeau, B. A., Comerford, M. P., Arakawa, C. K., Shadish, J. A. & Deforest, C. A. Engineered modular biomaterial logic gates for environmentally triggered therapeutic delivery. *Nat. Chem.* **10**, 251–258 (2018).
6. Kloxin, A. M., Kasko, A. M., Salinas, C. N. & Anseth, K. S. Photodegradable hydrogels for dynamic tuning of physical and chemical properties. *Science (80-.).* **324**, 59–63 (2009).
7. Rapp, T. L. & DeForest, C. A. Visible Light-Responsive Dynamic Biomaterials: Going Deeper and Triggering More. *Adv. Healthc. Mater.* **9**, 1–15 (2020).
8. Al-Afyouni, M. H., Rohrabaugh, T. N., Al-Afyouni, K. F. & Turro, C. New Ru(ii) photocages operative with near-IR light: New platform for drug delivery in the PDT window. *Chem. Sci.* **9**, 6711–6720 (2018).
9. Rapp, T. L. & DeForest, C. A. Targeting drug delivery with light: A highly focused approach. *Adv. Drug Deliv. Rev.* **171**, 94–107 (2021).
10. Lameijer, L. N. *et al.* A Red-Light-Activated Ruthenium-Caged NAMPT Inhibitor Remains Phototoxic in Hypoxic Cancer Cells. *Angew. Chemie - Int. Ed.* **56**, 11549–11553 (2017).
11. Kand, D. *et al.* Organelle-Targeted BODIPY Photocages: Visible-Light-Mediated Subcellular Photorelease. *Angew. Chemie - Int. Ed.* **58**, 4659–4663 (2019).
12. Peterson, J. A. *et al.* Family of BODIPY Photocages Cleaved by Single Photons of Visible/Near-Infrared Light. *J. Am. Chem. Soc.* **140**, 7343–7346 (2018).
13. Rubinstein, N., Liu, P., Miller, E. W. & Weinstain, R. Meso-Methylhydroxy BODIPY: A scaffold for photo-labile protecting groups. *Chem. Commun.* **51**, 6369–6372 (2015).
14. Kand, D. *et al.* Water-Soluble BODIPY Photocages with Tunable Cellular Localization. *J. Am. Chem. Soc.* **142**, 4970–4974 (2020).
15. Li, A., Turro, C. & Kodanko, J. J. Ru(II) polypyridyl complexes as photocages for bioactive compounds containing nitriles and aromatic heterocycles. *Chem. Commun.* **54**, 1280–1290 (2018).
16. Tu, Y. J. *et al.* Selective Photodissociation of Acetonitrile Ligands in Ruthenium Polypyridyl Complexes Studied by Density Functional Theory. *Inorg. Chem.* **54**, 8003–8011 (2015).
17. Silva, J. M., Silva, E. & Reis, R. L. Light-triggered release of photocaged therapeutics - Where are we now? *J. Control. Release* **298**, 154–176 (2019).
18. Aujard, I. *et al.* O-nitrobenzyl photolabile protecting groups with red-shifted absorption: Syntheses and uncaging cross-sections for one- And two-photon excitation. *Chem. - A Eur. J.* **12**, 6865–6879 (2006).

19. Thomas, S. W. New applications of photolabile nitrobenzyl groups in polymers. *Macromol. Chem. Phys.* **213**, 2443–2449 (2012).

20. Iturmendi, A., Theis, S., Maderegger, D., Monkowius, U. & Teasdale, I. Coumarin-Caged Polyphosphazenes with a Visible-Light Driven On-Demand Degradation. *Macromol. Rapid Commun.* **39**, 1–6 (2018).

21. Menge, C. & Heckel, A. Coumarin-caged dG for improved wavelength-selective uncaging of DNA. *Org. Lett.* **13**, 4620–4623 (2011).

22. Stowell, M. H. B., Rock, R. S., Rees, D. C. & Chan, S. I. Efficient synthesis of photolabile alkoxy benzoin protecting groups. *Tetrahedron Lett.* **37**, 307–310 (1996).

23. Salahi, F. *et al.* PHP-Tethered N-Acyl Carbamate: A Photocage for Nicotinamide. *Org. Lett.* **20**, 2547–2550 (2018).

24. Sun, W. *et al.* An Amphiphilic Ruthenium Polymetallodrug for Combined Photodynamic Therapy and Photochemotherapy In Vivo. *Adv. Mater.* **29**, (2017).

25. Zayat, L., Salierno, M. & Etchenique, R. Ruthenium(II) Bipyridyl Complexes as Photolabile Caging Groups for Amines. *Inorg. Chem.* **45**, 1728–1731 (2006).

26. Rapp, T. L., Wang, Y., Delessio, M. A., Gau, M. R. & Dmochowski, I. J. Designing photolabile ruthenium polypyridyl crosslinkers for hydrogel formation and multiplexed, visible-light degradation. *RSC Adv.* **9**, 4942–4947 (2019).

27. Rapp, T. L. & Dmochowski, I. J. *Ruthenium-cross-linked hydrogels for rapid, visible-light protein release*. *Methods in Enzymology* vol. 624 (Elsevier Inc., 2019).

28. Horbert, R., Pinchuk, B., Davies, P., Alessi, D. & Peifer, C. Photoactivatable Prodrugs of Antimelanoma Agent Vemurafenib. *ACS Chem. Biol.* **10**, 2099–2107 (2015).

29. Seio, K. *et al.* Photo-controlled binding of MutS to photo-caged DNA duplexes incorporating 4-O-(2-nitrobenzyl) or 4-O-[2-(2-nitrophenyl)propyl]thymidine. *Bioorganic Med. Chem. Lett.* **26**, 4861–4863 (2016).

30. Seyfried, P. *et al.* Optimal Destabilization of DNA Double Strands by Single-Nucleobase Caging. *Chem. - A Eur. J.* **24**, 17568–17576 (2018).

31. Höbartner, C. & Silverman, S. K. Modulation of RNA Tertiary Folding by Incorporation of Caged Nucleotides. *Angew. Chemie* **117**, 7471–7475 (2005).

32. Chen, M., Sun, W., Kretzschmann, A., Butt, H. J. & Wu, S. Nanostructured polymer assemblies stabilize photoactivatable anticancer ruthenium complexes under physiological conditions. *J. Inorg. Biochem.* **207**, (2020).

33. Josa-Culleré, L. & Llebaria, A. In the Search for Photocages Cleavable with Visible Light: An Overview of Recent Advances and Chemical Strategies. *ChemPhotoChem* 298–316 (2020) doi:10.1002/cptc.202000253.

34. Vorobev, A. Y. & Moskalensky, A. E. Long-wavelength photoremovable protecting groups: On the way to in vivo application. *Comput. Struct. Biotechnol. J.* **18**, 27–34 (2020).

35. Salassa, L., Garino, C., Salassa, G., Gobetto, R. & Nervi, C. Mechanism of Ligand Photodissociation in Photoactivable [Ru(bpy)2L2]2+ Complexes: A Density Functional Theory Study. *J. Am. Chem. Soc.* **130**, 9590–9597 (2008).

36. Wagenknecht, P. S. & Ford, P. C. Metal centered ligand field excited states: Their roles in the design and performance of transition metal based photochemical molecular devices. *Coord. Chem. Rev.* **255**, 591–616 (2011).

37. Havrylyuk, D., Stevens, K., Parkin, S. & C. Glazer, E. Toward Optimal Ru(II) Photocages: Balancing Photochemistry, Stability, and Biocompatibility Through Fine Tuning of Steric, Electronic, and Physiochemical Features. *Inorg. Chem.* **59**, 1006–1013 (2020).

38. Braslavsky, S. E. Glossary of terms used in photochemistry 3rd edition: (IUPAC Recommendations 2006). *Pure Appl. Chem.* **79**, 293–465 (2007).

39. Poblet, J. M., López, X. & Bo, C. Ab initio and DFT modelling of complex materials: Towards the understanding of electronic and magnetic properties of polyoxometalates. *Chem. Soc. Rev.* **32**, 297–308 (2003).

40. Ricci, G., San-Fabián, E., Olivier, Y. & Sancho-García, J. C. Singlet-Triplet Excited-State Inversion in Heptazine and Related Molecules: Assessment of TD-DFT and ab initio Methods. *ChemPhysChem* **22**, 553–560 (2021).

41. Venkatasubramanian, V. The promise of artificial intelligence in chemical engineering: Is it here, finally? *AIChE J.* **65**, 466–478 (2019).

42. Haghighatlari, M. & Hachmann, J. Advances of machine learning in molecular modeling and simulation. *Curr. Opin. Chem. Eng.* **23**, 51–57 (2019).

43. Hough, B. R., Beck, D. A. C., Schwartz, D. T. & Pfaendtner, J. Application of machine learning to pyrolysis reaction networks: Reducing model solution time to enable process optimization. *Comput. Chem. Eng.* **104**, 56–63 (2017).

44. Jiao, Z., Hu, P., Xu, H. & Wang, Q. Machine learning and deep learning in chemical health and safety: A systematic review of techniques and applications. *J. Chem. Heal. Saf.* **27**, 316–334 (2020).

45. Liu, K., Xin, H. & Han, M. Elucidation of key factors in nickel-diphosphines catalyzed isomerization of 2-methyl-3-butenenitrile. *J. Catal.* **377**, 13–19 (2019).

46. Wang, B., Zhou, L., Xu, K. & Wang, Q. Prediction of Minimum Ignition Energy from Molecular Structure Using Quantitative Structure–Property Relationship (QSPR) Models. *Ind. & Eng. Chem. Res.* **56**, 47–51 (2016).

47. Durant, J. L., Leland, B. A., Henry, D. R. & Nourse, J. G. Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci.* **42**, 1273–1280 (2002).

48. Nisbett, K., Tu, Y.-J., Turro, C., J. Kodanko, J. & Bernhard Schlegel, H. DFT Investigation of Ligand Photodissociation in [RuII(tpy)(bpy)(py)]2+ and [RuII(tpy)(Me2bpy)(py)]2+ Complexes. *Inorg. Chem.* **57**, 231–240 (2017).

49. Jacquemin, D., Planchat, A., Adamo, C. & Mennucci, B. TD-DFT assessment of functionals for optical 0-0 transitions in solvated dyes. *J. Chem. Theory Comput.* **8**, 2359–2372 (2012).