Rapid Prediction of Optical Absorption property for Tetrapyrrole-Based Photodynamic and Photothermal Therapy Agents

Guofan Zhang, M.S[1]; Rui Liu, PhD[1]; Fang-Fang Yin, PhD[1]
[1]Duke Kunshan University

2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

ABSTRACT

Purpose:
Photodynamic and photothermal therapies rely on tetrapyrrole photosensitizers with strong visible and nearinfrared absorption. However, accurate spectral prediction is often computationally expensive. This study develops a rapid machine learning framework to predict optical absorption properties for tetrapyrrolebased phototherapy agents.

Methods:
A dataset of 1,437 tetrapyrrole molecules, including porphyrins, chlorins, and bacteriochlorins, was constructed using density functional theory calculated frontier orbital energies. Molecular structures were encoded with SMILESderived fingerprints. Automated machine learning models were trained to predict HOMO-1, HOMO, LUMO, and LUMO+1 energies, which were then used to reconstruct absorption spectra.

Results:
The optimized models achieved R^2 values of 0.985 on validation data and 0.936 on independent test data, with mean absolute errors below 0.1 eV. Reconstructed spectra captured key Q-band and B-band features relevant to phototherapy activation.

Conclusion:
This framework provides a fast and accurate screening tool for designing new tetrapyrrolebased photodynamic and photothermal therapy agents.

CONTACT

Guofan Zhang
Duke Kunshan University
No. 8 Duke Avenue, Kunshan, Jiangsu Province, China 215316
gz83@duke.edu
gfzh.org

INTRODUCTION

Photodynamic therapy (PDT) and photothermal therapy (PTT) are minimally invasive lightbased cancer treatment strategies that depend on efficient optical activation of a photosensitizer or photothermal agent. In PDT, the activated molecule generates reactive oxygen species that damage tumor cells, while in PTT, absorbed light is converted into localized heat. For both approaches, strong absorption in the visible and nearinfrared region is essential because this spectral range supports deeper tissue penetration and more effective therapeutic activation.

Tetrapyrrole macrocycles, including porphyrins, chlorins, and bacteriochlorins, are important candidates for PDT and PDDDTT because their conjugated structures produce tunable absorption bands. Their optical behavior can be interpreted using Gouterman's fourorbital model, which links molecular structure to the energies of HOMO-1, HOMO, LUMO, and LUMO+1 orbitals. Structural modifications such as peripheral substitution, macrocycle saturation, and metal coordination can shift these frontier orbital energies and alter Q-band and B-band absorption features.

However, accurate prediction of absorption spectra using quantum chemical methods remains computationally expensive for large molecular libraries. This limits highthroughput screening of new phototherapy agents. To address this challenge, this work develops a physicsguided machine learning framework that uses molecular fingerprints and frontier orbital theory to rapidly predict optical absorption properties of tetrapyrrole compounds. By connecting molecularscale prediction with PDT/PTT activation mechanisms, the framework supports efficient screening and rational design of nextgeneration phototherapy drug candidates.

METHODS AND MATERIALS

A curated dataset of 1,437 tetrapyrrole macrocycles, including porphyrins, chlorins, and bacteriochlorins, was used to develop a physicsguided machine learning framework. For each molecule, density functional theory calculated HOMO-1, HOMO, LUMO, and LUMO+1 energies were collected based on Gouterman's fourorbital model.

Molecular structures were encoded using SMILESderived fingerprints and used as inputs for automated machine learning models. The predicted frontier orbital energies were then used to reconstruct UV-Vis-NIR absorption spectra, with emphasis on Q-band and B-band positions and coverage within the 650 to 800 nm therapeutic window. Model performance was evaluated using R^2, mean absolute error, root mean squared error, mean squared error, and mean relative error on validation and independent test sets.

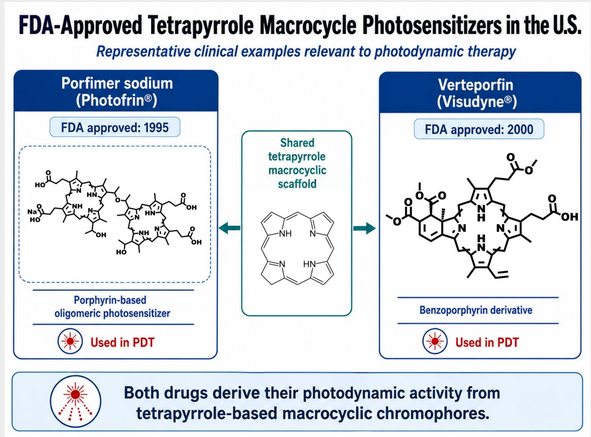

Figure 1. Representative FDAapproved tetrapyrrole macrocycle photosensitizers used in photodynamic therapy. Porfimer sodium (Photofrin®) and verteporfin (Visudyne®)

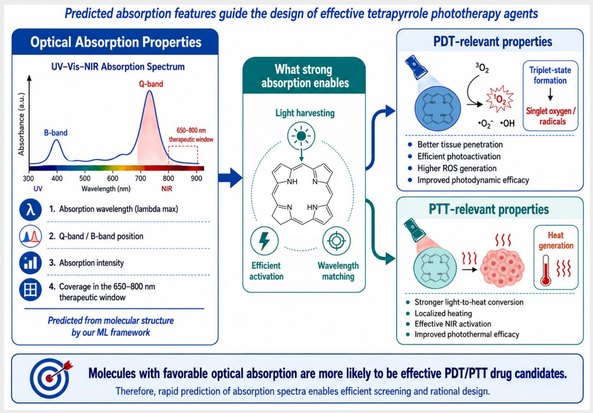

Figure 2. Optical Absorption Properties Govern PDT/PTT Drug Performance

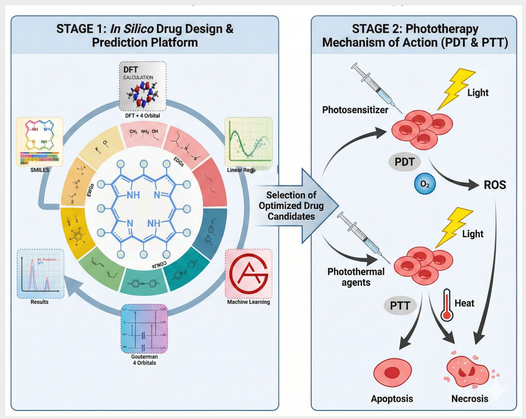

Figure 3. Integrated Workflow for Designing and Predicting Drugs for Photodynamic & Photothermal Therapy

RESULTS

The physicsguided machine learning framework accurately predicted frontier orbital energies and reconstructed optical absorption spectra for 1,437 tetrapyrrole macrocycles, including porphyrins, chlorins, and bacteriochlorins. Using SMILESderived molecular fingerprints, the automated learning models predicted HOMO-1, HOMO, LUMO, and LUMO+1 energies with strong performance. The optimized models **achieved R^2 values of 0.985 on the validation set and 0.936 on the independent test set, with mean absolute errors below 0.1 eV**. Reconstructed spectra captured key Q-band and B-band features relevant to phototherapy activation. Model performance remained consistent across the full tetrapyrrole database and chemically distinct molecular clusters.

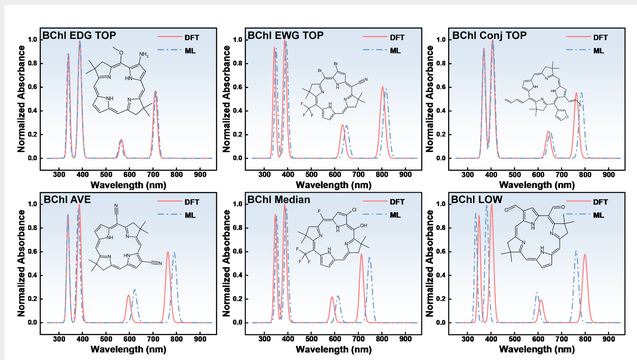

Figure 4. Our model's predictions are compared to the ground truth. TOP indicates the best performance, AVE represents the average, Median represents the median, and LOW indicates the worst performance.

DISCUSSION

These results demonstrate that optical absorption properties of tetrapyrrolebased phototherapy agents can be rapidly estimated from molecular structure without performing timeintensive excitedstate calculations for every candidate. Because Q-band position, absorption intensity, and therapeutic-window coverage are directly related to PDT and PTT efficacy, the predicted spectra provide useful criteria for candidate selection. The framework preserves physical interpretability by linking molecular fingerprints to Gouterman fourorbital frontier energies before reconstructing absorption spectra. This makes the model more transparent than a purely datadriven blackbox approach. The method is especially useful for highthroughput screening, where many candidate photosensitizers or photothermal agents must be prioritized before synthesis, DFT validation, or experimental testing.

CONCLUSIONS

This study presents a rapid, interpretable, and scalable machine learning framework for predicting optical absorption properties of tetrapyrrolebased PDT and PTT drug candidates. By combining molecular fingerprints, frontier orbital theory, and automated learning, the method accurately predicts key electronic features that determine phototherapy activation. The framework enables efficient screening of large tetrapyrrole libraries and supports rational design of nextgeneration photosensitizers and photothermal agents with improved absorption in clinically relevant wavelength regions.

REFERENCES

1. Karotki, Aliaksandr, et al. "Simultaneous two-photon excitation of photofrin in relation to photodynamic therapy." Photochemistry and photobiology 82.2 (2006): 443-452.
2. Verteporfin in Photodynamic Therapy Study Group. "Photodynamic therapy of subfoveal choroidal neovascularization in pathologic myopia with verteporfin. 1-year results of a randomized clinical trial--VIP report no. 1." Ophthalmology 108.5 (2001): 841-852.
3. Zhang, Guofan, et al. "A versatile method for accurately predicting electronic absorption spectra of tetrapyrrole macrocycles." Scientific Reports (2026).