CHEMFLUOR-VAE: REVERSE DESIGN OF ORGANIC FLUOROPHORES BASED ON EXPERIMENTAL OPTICAL PROPERTIES AND VARIATIONAL AUTOENCODER

CHENGWEI JU, *Pritzker School of Molecular Engineering, The University of Chicago, Chicago, IL, USA*; YONGRUI LUO, *Key Laboratory of Organofluorine Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences,, Shanghai, China*; BO LI, *Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, USA*; YUZHI XU, *Department of Chemistry, New York University, New York, NY, USA*; HANZHI BAI, *Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China*; RUIMING LIN, ZEHAN MI, HAOZHE ZHANG, *Pritzker School of Molecular Engineering, The University of Chicago, Chicago, IL, USA*.

Organic fluorescent molecules with desired optical properties attracted great attention, while the rational design was hindered by unclear structure properties relationship and the lack of rapid/affordable prediction methods. With the introduction of statistics-based methods in the prediction of photophysical properties for organic dyes, reverse design of fluorophores without traversing chemical space is still challenged by the features used for current methodologies.

In this work, we construct a self-referencing embedded strings (SELFIES)-based variational autoencoder (VAE) and a prediction model, which uses the latent space as the input, for the organic fluorophores, in the absence of joint training. The VAE can reproduce the structure of midsize organic dyes with acceptable accuracy. A tree-based prediction model based on Gradient Boosted Regression Trees (GBRT) can estimate the optical properties of organic dyes with a MAE 0.134 eV for emission energy and an accuracy of 0.81 for photoluminescence quantum yield (PLQY), which is comparable with the state-of-the-art quantum-mechanical based approach, time-dependent density-functional theory (TD-DFT). The feasibility of our approach in reverse design is proved by preliminary attempts at skeleton optimization and validated by first-principles calculations. New experimental synthesized molecules demonstrated the accuracy of our prediction model. Meanwhile, due to the continuous values in the latent space, this VAE-based methodology makes gradient optimization become possible for large organic materials. Combined, our statistical learning methodology opens a new venue for the design of organic fluorophore, can also be extended to the field of organic solar cell (photo conversion efficiency, PCE) and organic field-effect transistor (conductivity).