

# A Computational Framework for the Efficient Design of Organic Dyes with Targeted Optical Properties

Mariia Kyrpa<sup>1</sup>, Volodymyr V. Ivanov<sup>1</sup>

<sup>1</sup>Physical Chemistry and Materials Chemistry Department, School of Chemistry, V. N. Karazin Kharkiv National University, 4 Svobody sq., 61022, Kharkiv, Ukraine

An important branch of modern chemistry and materials science is the study of the spectral properties of organic dyes, represented by strongly polar  $\pi$ -conjugated systems, since they find extremely wide and diverse applications in engineering, optics, photonics, molecular electronics, in particular due to the notable solvatochromism and nonlinear optical properties. Such molecules are used as probes for surface studies, biological markers, dyes in the food industry, etc. By means of chemoinformatics combined with quantum chemistry, it is possible not only to qualitatively describe the spectral behaviour of potential dyes, but also to identify the parameters that determine it the most. Nowadays, such studies are inevitably associated with working with large libraries of molecules, so the development of methods that would give a satisfactory result with minimal computer costs remains relevant.

As a background to this study, we have validated the DFTB (Tight-Binding Density Functional Theory) method [1] – an accelerated and computationally effective alternative to DFT, based on approximating some integrals by use of pre-parameterized values to describe the interactions between pairs of atoms. On the example of the Effenberger dye [2] (figure 1) isomers library [3], consisting of 59 molecules, it was shown that the DFTB method is capable of reproducing the main trends in the spectral behaviour for given set of molecules, in comparison with the reference level of theory, DFT [4]. We have also proposed a hybrid approach, consisting in a combination of DFT and DFTB, with further reproducibility improvement, but still requiring expensive calculations.

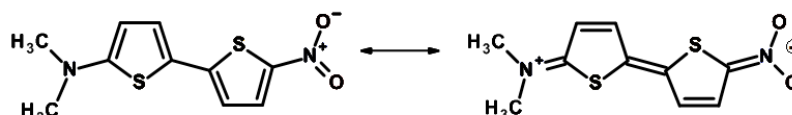


Figure 1. Resonance structures of Effenberger Dye.

In order to increase the screening efficiency, we trained the artificial neural network (ANN) based on the electronic transition energy values obtained from DFTB, as well as a number of other descriptors, such as the HOMO-LUMO gap energy, non-planarity and sphericity indices, etc. For compact representation of the electronic density spatial distribution we developed a set of novel orbital descriptors derived from the orbital shape tensor and its eigenvalues, such as effective orbital size, dimensionality index and effective orbital volume.

In addition, as a basic model, we developed a regression equation for the descriptors that are most correlated with the described property. The obtained coefficient of determination  $R^2$  and the coefficient of determination in accordance with Leave-One-Out Cross-Validation are 0.680 and 0.645, respectively. The use of a ANN significantly enhanced the model's predictive capabilities: the obtained  $R^2$  values are 0.814 for the training set and 0.753 for the test set.

We expect that this approach will be transferred to larger and more structurally diverse libraries, and will also be used to predict other characteristic properties of organic dyes.

[1] R. Rüger; E. Van Lenthe; T. Heine; L. Visscher. *J. Chem. Phys.* 144 (2016) 184103.

[2] F. Effenberger; F. Wuerthner; F. Steybe. *J. Org. Chem.* 60 (1995) 2082-2091.

[3] A. B. Zakharov; M. Kyrpa; A. V. Kyrychenko; S. M. Kovalenko; O. N. Kalugin; V. V. Ivanov; L. Adamowicz. *Struct. Chem.* 36 (2025) 723-738.

[4] M. Kyrpa; A. Zakharov; V. Ivanov. *Kharkiv Univ. Bull., Chem. Ser.* 44 (2025) 72-80.