

# Prediction of absorption wavelengths using a combination of semi-empirical quantum mechanics simulations and quantitative structure–property relationship modeling approaches



R. Chaudret\*, Cs. F. Kiss, L. Subramanian

Scienomics, 16 rue de l'Arcade, 75008 Paris, France

## ARTICLE INFO

### Article history:

Received 16 September 2014  
Received in revised form 19 November 2014  
Accepted 22 November 2014  
Available online 24 November 2014

### Keywords:

QSPR  
Semi-empirical calculations  
Polycyclic aromatic molecules  
OLED  
UV absorption prediction

## ABSTRACT

Light emitting or absorbing molecules have a wide range of industrial applications. To enable *in-silico* screening of a number of such molecules one should be able to predict their emission or absorption spectra rapidly and accurately. In the present study, we developed a combination of semi-empirical quantum mechanical (QM) simulations and a quantitative structure–property relationship (QSPR) modeling approach to compute such properties rapidly with a fair degree of accuracy. We focused our work on the prediction of the first absorption wavelength of a set of 40 fluorophores or polycyclic aromatic molecules. We also compared the results obtained from our QM–QSPR combination approach with those from different stand-alone semi empirical QM calculations and pure QSPR modeling approach. We showed that there is an advantage in the QM–QSPR combination approach for predicting and screening optical properties quickly and with a great deal of accuracy.

© 2014 Elsevier B.V. All rights reserved.

## 1. Introduction

Light emitting small molecules or polymers have a wide range of applications from chemical sensors in analytical chemistry and biochemistry to OLED (organic light emitting diode) devices used in television, computer monitors or mobile phones. Their potential for low cost of production and high light emission specificity make them very promising materials. Accurate prediction of molecular UV absorption spectra is crucial in order to propose, design and develop new molecules. Since the development cost and time of a new molecule can often be rather large, a low cost and quick screening procedure is required to provide interesting initial guesses for focused experimental work in a smaller and more promising design space. This low cost and quick screening approach should, however, fulfill three major requirements:

- The approach should be rapid for each molecule.
- The approach should be automated in order to test many different molecules.
- The approach should be reliable and the accuracy should be known even if it is not very high in the first round of screening, to

provide direction towards the design of the best potential molecules.

In the past years, many different *ab-initio* quantum mechanical (QM) methods have been developed in order to predict accurately UV absorption/emission spectra of organic molecules.

QM approaches such as CASSCF, CASPT2, MR-PT2 or DFT based methods (TD-DFT) were shown to reach a very good accuracy in the prediction of excitation energies (and therefore adsorption wavelengths) [1–6]. However, such approaches remain very time consuming and are therefore limited to small and middle size molecules. In order to decrease the computational effort of the calculation, semi-empirical approaches were largely developed. Traditional MNDO-like approaches are very efficient but tend to underestimate the energy of excited states and therefore predict longer absorption wavelengths. A modified version of INDO [7] (Intermediate Neglect of Differential Orbital), INDO/S [8], that take into account  $\pi$ -type overlap has long been a reference method for excited-state calculations. In order to better reproduce spectral data, Ridley and Zinder proposed another modified version named ZINDO/S [9]. It was shown that this approach, as well as well-known PM3 [10] method, allowed a good representation of  $\pi \rightarrow \pi^*$  type transitions [11] (that will be studied here). Alternatively the linear combination of orthogonalized atomic orbitals (LCAO) scheme proposed by Spengert-Larsen [12,13] was shown to predict with a good accuracy the spectra of polycyclic aromatic

\* Corresponding author. Tel.: +33 140175672.

E-mail address: [robin.chaudret@scienomics.com](mailto:robin.chaudret@scienomics.com) (R. Chaudret).

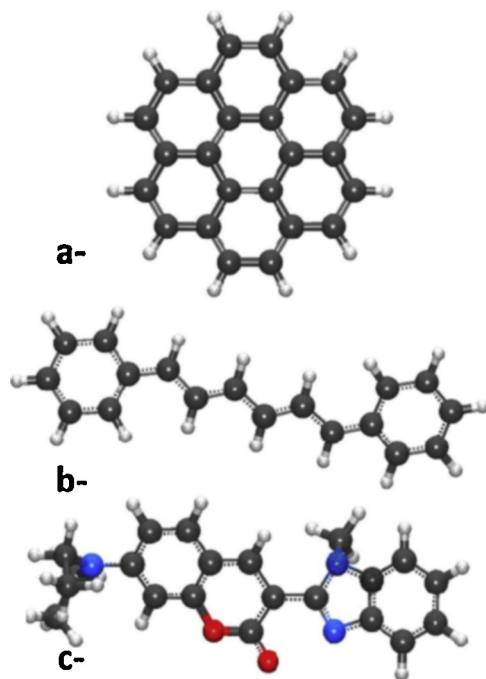
hydrocarbon molecules [14–17]. Finally, orthogonalization models (OMx) were developed to include the effects of Pauli exchange repulsion [18]. Such approaches were proved to reach a very good accuracy in the prediction of UV spectra. In 2010 different semi-empirical methods (7 different semi-empirical approaches) were tested over a set of 28 medium size organic molecules [19]. This study showed that the OMx methods were giving the lowest deviation for the calculation of both singlet and triplet excitation.

In the past decade, QSPR approaches have also been used to model molecular UV–spectral properties. For example Katritzky group managed to model the absorbance [20] ( $R^2=0.815$ ) or intensities [21] (0.692) of a large set of compounds. Liu et al. also obtained interesting results on the prediction of maximum absorption wavelength but on more specific group of molecule such as derivatives of coumarins [22] or flavones [23] molecules.

In this study we propose to leverage the advantages of both semi-empirical QM simulation approach and quantitative structure property relationship (QSPR) modeling approach to improve the predictions without increasing the calculation time. In order to perform accurate screening of potentially interesting chromophores or OLED molecules the model was developed on a large set of aromatic molecules.

## 2. Method

All work reported here were performed using MAPS 3.4 platform [24], with the modules called MNDO and QSAR within MAPS. 40 organic molecules taken from the fluorophore and OLED Sigma–Aldrich databases were sketched within MAPS (see Fig. 1 for a representation of coronene, 1,6-diphenylhexatriene and coumarin 30 molecules). All the molecules selected have a strong aromatic delocalization. For this set of molecules the absorption wavelength ranges between 247 nm and 699 nm and are the band absorption maxima. These experimental values were extracted from the fluorophore [25] and Jinno [26] databases.



**Fig. 1.** Representation of some of the molecules studied: (a-) coronene, (b-) 1,6-diphenylhexatriene and (c-) coumarin 30. Carbon atoms are shown in black, hydrogen in white, nitrogen in blue and oxygen in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

For this study only molecules with experimental spectra obtained in solvents interacting weakly with the molecules were selected (such as cyclohexane or tetrachloromethane). This was done to achieve a consistency of the data with regard to the solvent effects. Therefore, in the following, the effects of solvent on the absorption spectra will be neglected.

The sketched molecules were first rapidly relaxed using UFF force field [27] available in the MAPS platform. The obtained structures were then optimized at semi-empirical level using the module available in MAPS. Since the goal of this study is not to perform a benchmark of different semi-empirical methods for the calculation of the absorption spectrum of polycyclic aromatic hydrocarbons, the widely used methods MNDO [28] and PM3 [29] methods were selected in this work. In addition a recent approach named OM3 [30] was also tested. PM3 was found to have a good accuracy for modeling  $\pi \rightarrow \pi^*$  type transitions [11] in a benchmark test of 21 molecules and may, therefore, reproduce well the excitation energies for the molecules in this study. Similarly OM3 method appeared to be the most accurate semi-empirical approach in a set of 7 methods tested) for the reproduction of excitation energies of 28 middle size organic molecules.

For the optimized structures, the absorption spectra were calculated using the GUGA-CI [31,32] method with 5 occupied and 5 virtual orbitals involved in the active space and with calculation of single, double and triple excitations (CISDT). Finally the first 10 excited states were computed. In order to facilitate the calculation, a script was created using the MAPS Python interface to automatically start the calculations for all 40 molecules. The first absorption wavelength was then extracted from MNDO output as explained in Supplementary information and stored into MAPS database in order to make them easily available for their use for quantitative structure–property relationship modeling. In the following of the study, the PM3 optimized structure will be used for all molecules.

Quantitative structure property relationship (QSPR) approach considers that a given molecular property (whether it is drug activity, liquid viscosity, absorption wavelength, etc.) can be modeled using finite set of molecular “features” called descriptors. These “features” can be related to the composition (for example number of carbon or hydrogen atoms), topology (number of rings, double or single bonds) or structure (distance between atoms, angles, torsions . . . ) of the molecule. This list is obviously not extensive and many other molecular properties can also be used. The goal of a QSPR analysis is then to create a regression model between the different descriptors and the property of interest. In order to find the most appropriate descriptors, MAPS QSAR module allows to compute 4885 Dragon molecular descriptors [33] but also to utilize user-defined descriptors. The program also contains several methods to reduce the number of considered descriptors. Using this module in MAPS, two QSPR models were regressed: (i) one using only Dragon descriptors and the other (ii) using the PM3 computed wavelengths together with the Dragon descriptors.

First, Dragon descriptors from different families (constitutional indices, ring descriptors, topological indices, geometry descriptors, atom-centered fragments and molecular descriptors) were calculated. After removing zero-variance descriptors, over 150 descriptors were still remaining. We used the MAPS QSAR module to remove highly correlated descriptors (correlation threshold=0.85), i.e., those having similar effects on the current molecular set. This procedure reduced the number of descriptors to around 40. These 40 descriptors were then used as starting point for the creation of models. Both genetic algorithm and model refinement approaches were used to reduce drastically the number of descriptors for the final models. This resulted in 4 or 5 descriptors maximum at the final stage.

Download English Version:

<https://daneshyari.com/en/article/26410>

Download Persian Version:

<https://daneshyari.com/article/26410>

[Daneshyari.com](https://daneshyari.com)