



Prediction of the maximum absorption wavelength of azobenzene dyes by QSPR tools

Xuan Xu^a, Feng Luan^{a,*}, Huitao Liu^a, Jianbo Cheng^a, Xiaoyun Zhang^b

^a Department of Applied Chemistry, Yantai University, Yantai 264005, PR China

^b Department of Chemistry, Lanzhou University, Lanzhou 730000, PR China

ARTICLE INFO

Article history:

Received 22 March 2011

Received in revised form 1 July 2011

Accepted 24 August 2011

Keywords:

Azobenzene dyes

Maximum absorption wavelength

QSPR

MLR

RBFNN

Molecular modeling

ABSTRACT

The maximum absorption wavelength (λ_{\max}) of a large data set of 191 azobenzene dyes was predicted by quantitative structure–property relationship (QSPR) tools. The λ_{\max} was correlated with the 4 molecular descriptors calculated from the structure of the dyes alone. The multiple linear regression method (MLR) and the non-linear radial basis function neural network (RBFNN) method were applied to develop the models. The statistical parameters provided by the MLR model were $R^2 = 0.893$, $R_{\text{adj}}^2 = 0.893$, $q_{\text{LOO}}^2 = 0.884$, $F = 1214.871$, $RMS = 11.6430$ for the training set; and $R^2 = 0.849$, $R_{\text{adj}}^2 = 0.845$, $q_{\text{ext}}^2 = 0.846$, $F = 207.812$, $RMS = 14.0919$ for the external test set. The RBFNN model gave even improved statistical results: $R^2 = 0.920$, $R_{\text{adj}}^2 = 0.919$, $q_{\text{LOO}}^2 = 0.898$, $F = 1664.074$, $RMS = 9.9215$ for the training set, and $R^2 = 0.895$, $R_{\text{adj}}^2 = 0.892$, $q_{\text{ext}}^2 = 0.895$, $F = 314.256$, $RMS = 11.6427$ for the external test set. This theoretical method provides a simple, precise and an alternative method to obtain λ_{\max} of azobenzene dyes.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

The history of dyes can be traced back to 2600 BC when there was the earliest written record of the use of dyes in China [1]. Originally, the dyes were obtained naturally until the discovery of the first synthetic dye pigment “Mauve” by William Henry Perkin in 1856 [2]. Since the first azo dye spliced onto the fabric by coupling in 1880, thousands of organic chromogens were produced by the dye industry [3]. For example, in 1995, 90% of the 3000 compounds registered in the Color Index were used at the level of at least 100 tons per year [4]. Nowadays, the abundant classes of colored organic compounds play a crucial role in the industry, not only as dyes or pigments but also in more technological fields, such as thermal transfer systems, molecular switches, media storages or photovoltaic devices [5–7]. Among so many kinds of dyestuffs, two families of industrial organic dyes, 9,10-anthraquinones (AQ) and azobenzenes (AB, see Fig. 1), encompass about 90% of today’s world dye production together [6]. The latter (AB) are also of importance in many branches of chemistry, being widely applied as dyes with colors ranging from red to blue [8]. As it is known, the industrial applications imply that it should design specific dyes possessing given properties including acidity, alkalinity, thermal stability, affinity to fibre, and adsorption. Another spectroscopic property, which is

related to the important color property and determined mainly by the structure of the dyes, is the maximum absorption wavelengths (λ_{\max}). Once the λ_{\max} of the compound is known, the color will be unequivocally determined.

Nowadays, the λ_{\max} of the dyes is usually obtained experimentally [9]. Application of experimental methods to obtain the λ_{\max} of dyes is the most obvious and effective methods, however, there are some drawbacks such as the need for laboratory facilities, the huge workload. Also, the method cannot be easily applied for toxic, volatile, explosive or radioactive substances; and it cannot be used if the material has not been synthesized yet. For instance, several classes of dyes are considered as possible carcinogens or mutagens; the high coloring power of dyes gives rise to esthetic damage: dye concentrations lower than 1 mg/L may induce visible coloration and hence public complaint [10]. Therefore, it is necessary to resort to the theoretical methods to compensate the shortage of experimental research methods. Several researchers have addressed the challenging task of predicting the λ_{\max} by the means of computation. We think that the methods for quantitative prediction of the absorption maxima of dyes from their molecular structures alone would be of significant utility not only in the use of dyes but also in the molecular design process of new dyes.

Some theoretical methods have been applied to solve this problem. One kind of theoretical method is the large-scale highly correlated ab initio approaches such as EOM-CC, MR-CI, or CAS-PT2, which can be used to calculate the maximum absorption wavelength of dyes based on post-Hartree–Fock methods [11–13].

* Corresponding author. Tel.: +86 535 6902063; fax: +86 535 6902063.

E-mail address: fluan@sina.com (F. Luan).

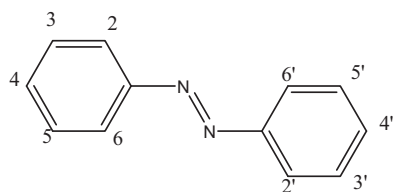


Fig. 1. General sketch of azobenzenes (AB) investigated in the present study.

Though they are accurate but time consuming and present a problematic scaling with the system dimension. On the other hand, semi-empirical methods (CNDO/S, INDO/S) are applicable for large systems but with a significant loss in accuracy [14,15]. In the recent years, the most widely applied tool for modeling electronic spectra of structures is the time dependent density functional theory (TD-DFT) [16–18], which can give a counterpoise between the remarkable accuracies and “reasonable” computational times [19–23]. It is important to stress that TD-DFT methods – although expensive for large systems – have been proven to be very efficient to calculate the λ_{\max} of organic and inorganic dyes including the azobenzene molecules [24–27].

Another theoretical method, fast quantitative structure–property relationship (QSPR) approaches, would possess some advantages, because they could be extended to larger molecules consisting of hundreds of atoms with acceptable accuracy and could deal with larger data sets quickly. In the past decades, QSPR studies, as one of the most important areas in chemometrics, were used widely to predict a wide variety of physical, chemical, biological and toxicology properties [28,29]. The purpose of this method is to seek for mathematical equations relating the descriptors calculated from the chemical structure to the desired properties [30]. As we know, there are not only over thousands of molecular descriptors to represent the attributes of the interesting compounds but also some of the descriptors always can correlate to the property properly [31]. The advantage of this approach lies in the fact that it requires only the knowledge of the chemical structure and is not dependent on any experimental properties. Once a correlation is established, it can be applied for the prediction of the property of new compounds that have not been tested, synthesized or found yet. Thus the QSPR approach can expedite the process of the development of new molecules and materials with desired properties.

There are some reports about the applications of QSPR methods to investigate the relationship between the λ_{\max} and the structure of dyes in recent years. For example, Buttingsrud et al. [32] build empirical models relating bond lengths and critical points in the electron density distribution to the λ_{\max} of azobenzene dyes. Xu et al. developed a linear QSPR model to predict maximum absorption wavelength of second-order NLO chromophores [33]. The same authors also predicted the maximum absorption wavelength of dye-sensitized solar cells using two different QSPR methods [34,35]. Li et al. applied semiempirical quantum chemical method and artificial neural networks to compute λ_{\max} of 31 azo dyes [36]. Moreover, a recent work of Fayet et al. performed a well QSPR study of the relation between the experimental λ_{\max} of a small set of 22 azobenzenes and 24 anthraquinones dyes using quantum chemical descriptors [37].

To the best of our knowledge, there has been no general QSPR study on the kind of azobenzenes dyes using the general descriptors. Together with the desire of introducing this method into dyes studies, in the present work, we aim to seek for a fast and general approach that would offer reliable theoretical prediction of λ_{\max} of the commonly used azobenzenes dyes.

2. Materials and methods

2.1. Data set

Our models were developed using a library of 191 azobenzene dyes. The data set was available from the literature published by Buttingsrud et al. [32]. The molecules used in this study, all have the common azobenzene skeleton shown in Fig. 1. The detail structures of all the studied molecules are shown in Supplementary Information (see S1). The experimental λ_{\max} of the azobenzene dyes and their corresponding number are listed in Table 1. These values cover the range from 318 to 514 nm, therefore within the ultraviolet and the visible range.

As it is usually done in any QSPR study, the entire set of compounds was randomly divided into the training and the external test sets. The training set, consisting of 152 compounds, was used to build and consider the robustness of the model. The test set with the other 39 compounds not used in the training set was used to evaluate the models once they were built.

2.2. Descriptor generation

The structures of all compounds were first drawn with the aid of the ISIS Draw 2.3 program [38]. Reasonable starting geometries were obtained by resorting to the MM+ molecular mechanics force field [39,40] in the HyperChem6.0 program [41]. The final optimization was obtained using the semi-empirical PM3 model implemented in the MOPAC 6.0 program [42]. All calculations were carried out at the restricted Hartree–Fock level (therefore with no configuration interaction). The molecular structures were optimized using the Polake–Ribiere algorithm until the root-mean-square gradient was less than 0.01 kcal/mol.

Subsequently, the output files exported from MOPAC were transferred into the software CODESSA [43,44] to generate the descriptors. In the present study, we obtained 38 constitutional, 38 topological, 12 geometrical, 81 electrostatic, and 396 quantum-chemical descriptors for each compound. In addition, 7 physicochemical descriptors were also calculated by Hyperchem, including refractivity, approximate surface area, grid surface area, volume, log *P*, polarizability and mass [41]. At last, 572 descriptors for each compound were kept for further study.

2.3. Modeling techniques

Generally, a huge number of descriptors representing a compound can be measured or computed, but how to deal with this high dimensional information is a problem. In many cases, there is no *a priori* knowledge on their role in describing a particular property. Therefore, it is very important to select a proper subset of descriptors to build the QSAR or QSPR models. Up to now, several methods used to select the descriptors have been reported, including linear and nonlinear methods like the multiple linear regression (MLR) [45], linear discriminant analysis (LDA) [46], principal component analysis (PCA) [47], and the global search methods like genetic algorithms (GA) [48] and simulated annealing (SA) [49], also including artificial neural network (ANN) [50] and support vector machines (SVMs) [51].

In the present study, the heuristic method (HM) was used. HM is implemented in CODESSA and has been used in several other papers [52–55], as it is 2–5 times faster than other methods with comparable quality [56]. The HM of the descriptor selection proceeds with a pre-selection by eliminating descriptors that (i) are not available for each structure; (ii) have a small variation in magnitude for all structures; (iii) have a Fisher *F*-criterion below 1.0; and (iv) have *t*-values less than the user-specified value (by default

Download English Version:

<https://daneshyari.com/en/article/1235191>

Download Persian Version:

<https://daneshyari.com/article/1235191>

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