QSPR modeling of UV absorption intensities

**Alan R. Katritzky · Svetoslav H. Slavov ·
Dimitar A. Dobchev · Mati Karelson**

Abstract Literature UV absorption intensities at 260 nm and 25 °C in water of a diverse set of 805 organic compounds when analyzed by CODESSA Pro software using an initial pool of 800 + descriptors provide a significant QSPR correlation ($R^2 = 0.692$). Concurrently, a neural networks approach was used to develop a corresponding nonlinear model. The descriptors appearing in these models are discussed with respect to the physical nature of the UV absorption phenomenon.

Introduction

High performance liquid chromatography (HPLC) combined with ultraviolet (UV) spectrophotometric detection is the method most applied in organic chemistry for analyzing reaction products [1].

UV is also considered a nearly universal detector for drug-like molecules: 85% of the structures in the MDDR (a database of drugs and candidate drugs [2]) contain an aromatic group and most of the remaining 15% contain another chromophore. A computational method for the prediction of the relative response of organic molecules in the UV region of spectra would be useful for scientists in this field, which is remarkably unexplored. While sets of derivatized molecules based on a common chromophore have been used for quantitation, [3], we located only two reports of attempts to treat UV intensities with QSPR: an earlier study from our group [1] and a treatment limited to polychlorinated biphenyls [4].

In recent years, many researchers have addressed the challenging task of predicting electronic absorption parameters by quantum theory [5–14]. One successful model for the calculation of UV spectra has been the ZINDO modification [15] of INDO (Intermediate Neglect of Differential Overlap). The so-called ZINDO/S is a modified intermediate neglect of differential overlap (INDO) method parameterized to reproduce UV/visible spectral transitions. New adjustable empirical parameters (σ–σ and π–π overlap weighting factors) were introduced into ZINDO/S to modify the resonance integrals for the off-diagonal elements of the Fock matrix. ZINDO predicts with high precision the experimental UV-vis spectra of the transition metal complexes for which it was specifically parameterized. ZINDO also works well for many other organic systems with extended conjugation. However, for systems containing nonbonded electron pairs ZINDO/S reproduces the experimental data poorly [16].

More recently *ab initio* predictions of UV spectra were also carried out by using highly correlated methods such as: Configuration Interaction Singles (CIS) or Time Dependent – Density Functional Theory (TD-DFT) [17] combined with high order basis sets: cc-pVTZ + sp and respectively B3LYP [18], but these calculations require powerful computing capacity and are currently time

A. R. Katritzky (✉) · S. H. Slavov · D. A. Dobchev
Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, Gainesville, FL 32611, USA
e-mail: katritzky@chem.ufl.edu

D. A. Dobchev · M. Karelson
Department of Chemistry, University of Tartu, 2 Jakobi Street, Tartu 51014, Estonia

S. H. Slavov · M. Karelson
Department of Chemistry, Tallinn University of Technology, Ehitajate tee 5, Tallinn 19086, Estonia

consuming for relatively large molecules. Therefore, fast QSPR approaches which could be extended to larger molecules consisting of hundreds of atoms with acceptable accuracy would possess significant advantages. In the present work, we sought a fast and general approach that would offer reliable theoretical prediction of UV spectral intensities by using CODESSA Pro [19] to obtain statistically significant and reliable QSPR models.

We are well aware that attempted correlation of ultraviolet spectra intensities in terms of extinction coefficients at a certain wavelength has no sound basis in theory as such coefficients depend in complex ways on the location and intensities of spectral maxima as well as the actual shape of the spectra. Thus it would undoubtedly be more correct to attempt a correlation of integrated intensities with theoretical descriptors. We have chosen the less sound approach for two main reasons. First, much more data is available for use as a training set if intensities at a certain wavelength are used. Secondly, and critically, all commercial HPLC UV detectors measure intensity at a single wavelength; to be able to use our model in practical applications we need to correlate the intensities at a constant wavelength.

CODESSA Pro has previously successfully correlated and predicted numerous physical and biological properties [20–24]. Hence, based on our experience, in this work along with the multilinear modeling we have also applied the backpropagation feed-forward artificial neural networks in order to obtain nonlinear QSPR models. We now report models for the prediction of the UV absorbance at 260 nm for a diverse set of 805 organic compounds using these quantitative structure-property relationship (QSPR) approaches.

Dataset

The NIST Chemistry WebBook [25] contains details of the UV/Vis spectra for over 800 compounds (http://webbook.nist.gov/chemistry/). The UV/Vis spectra data provided in this database were taken either from original scientific papers or from several published collections. All the UV/Vis values now utilized had been measured in aqueous solution. The JCAMP-DX file format was used to extract the logarithmic UV intensity values for 805 organic compounds. A linear extrapolation procedure was used to estimate the absorption values (in Log units) at 260 nm, as usually used in HPLC.

QSPR methodology

The three-dimensional conversion and pre-optimization of molecules were performed using the molecular mechanics force field MM+ as implemented in the HyperChem 7.5 package [26].

The final geometry optimization of the molecules was carried out using the semi-empirical quantum-mechanical AM1 parameterization [27]. The optimized geometries were loaded into CODESSA Pro software. Overall, more than 800 theoretical descriptors were calculated. These descriptors are generally classified into the following groups: (i) constitutional, (ii) topological, (iii) geometrical, (iv) thermodynamic, (v) quantum chemical and (vi) charge-related descriptors.

In the next step, a modified QSPR linear modeling approach was used. This aimed to combine the advantages of the two most frequently applied QSPR methods i.e. (i) use of all data points to build the model and to apply internal crossvalidation procedures for validation or (ii) to use only a part of the available data to build the model and the remaining data points for external validation. Our procedure to build a QSPR model was as follows:

1. All data points were arranged in ascending order of the experimental logUV intensity values.
2. The data set was separated into three subsets (conditionally denoted as A, B and C) collected by selection of every third point from the original dataset (from point 1) so that we obtained a similar distribution of the investigated property values (Fig. 1 and 2).
3. Three new datasets were constructed using the three binary sum combinations: A + B, A + C and B + C.
4. The standard QSPR modeling procedure encoded in CODESSA Pro as the ''Best Multiple Linear Regression Method'' (BMLR) [28, 29] was applied to the three datasets obtained in step 3 to derive three predictive models. In the BMLR treatment, the multilinear regression analysis commences with the set of two descriptors with pair $R^2_{ij} < 0.05$, i.e. with all the orthogonal or nearly orthogonal pairs of descriptors. The best two-parameter correlation from this set is found simply by performing the treatment of the property with each of the pairs. In the correlation treatment of the third rank (i.e. with three independent variables involved), each of the orthogonal descriptor pairs discussed above is combined with each of the other non-collinear ($R^2_{ij} < 0.5$) descriptors, and again the best correlation is chosen. Analogously, in the treatment of the forth and higher ranks, the 1,000 best variable sets for the previous rank are considered in combination with each of the additional non-collinear descriptor scales. To speed up the calculations all descriptors with insignificant variance are rejected.
5. External validation of the datasets A + B, A + C and B + C was carried out using the complementary parts to each (i.e. respectively C, B and A).
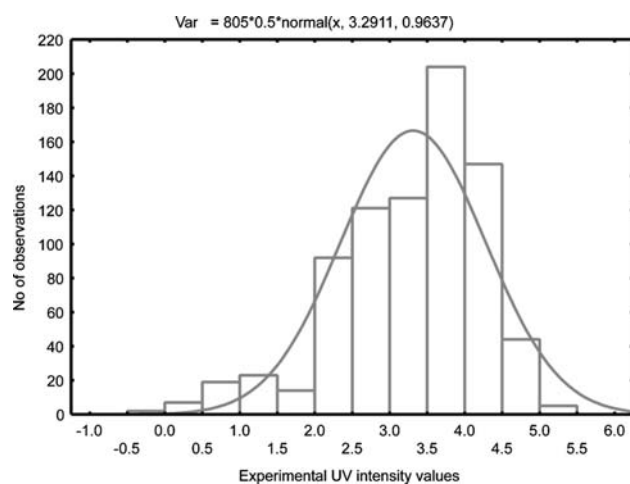
Fig. 1 The data distribution for the general population

6.　The total number of descriptors that appeared in each of the models obtained in step 4 for sets A + B, A + C and B + C were tested to obtain the best general model including all 805 compounds.

7.　The general model was again validated using classical leave-one-out and leave-many-out internal crossvalidation procedures.

The first four steps were also applied in our ANN modeling process [30] replacing BMLR procedure by the ''back propagation learning algorithm'' [31, 32]. Our ANN modeling also omitted the crossvalidation procedure given in step 7 due to the nonlinear nature of the ANN algorithm. However, the obligatory external validation was also performed for the neural networks modeling.

We investigated several different architectures for the ANN modeling. To compare the ANN results with these obtained by using of linear QSPR approach, the same descriptors as in the corresponding linear models were selected. The number of the neurons in the hidden layer was chosen so that the statistical parameters for the training and test sets would be as close as possible to each other to enhance the predictive power for the model.

The ANN network architecture finally chosen was 5–10–1, which means that the model had five input neurons in the input layer (representing the used descriptors), 10 hidden neurons in the hidden layer and one neuron in the output representing the predicted UV intensities values.

Results and discussion

For successful QSPR modeling, the data investigated should posses a normal distribution; furthermore the statistical parameters (mean, standard deviation and skewness) for the general population and for the samples should have similar magnitudes. As can be seen from the histograms shown in Fig. 1 for the general population and Fig. 2 for the samples, the data indeed posses a normal distribution and the mean, standard deviation and skewness correspond satisfactorily.

By applying the BMLR method the best QSPR models for each subset were derived. In our case the application of the ''breaking rule'' for defining the number of the descriptors in the equations (considering ΔR^2 values for the consecutive models with increased by one number of descriptors) did not provide an unambiguous stopping criterion. In this situation five descriptor linear models for A + B, A + C and B + C subsets were chosen (see Tables 1–3) based on the physicochemical meaning of the phenomena and its interpretation in terms of the calculated quantum-chemical descriptors. In Tables 1–4 $R^2_{cv}OO$ denotes the square of the leave-one-out cross-validated correlation coefficient; $R^2_{cv}MO$ is the square of the leave-many-out cross-validated correlation coefficient; RMSPEOO and RMSPEMO are calculated root mean square predictive errors in the case of leave-one-out and leave-many-out procedure, respectively. Additional statistical parameters used are summarized in Table 5.

To be able to compare the results obtained by the BMLR and ANN methods, we used as far as possible the same methodology steps and descriptors (as discussed in the
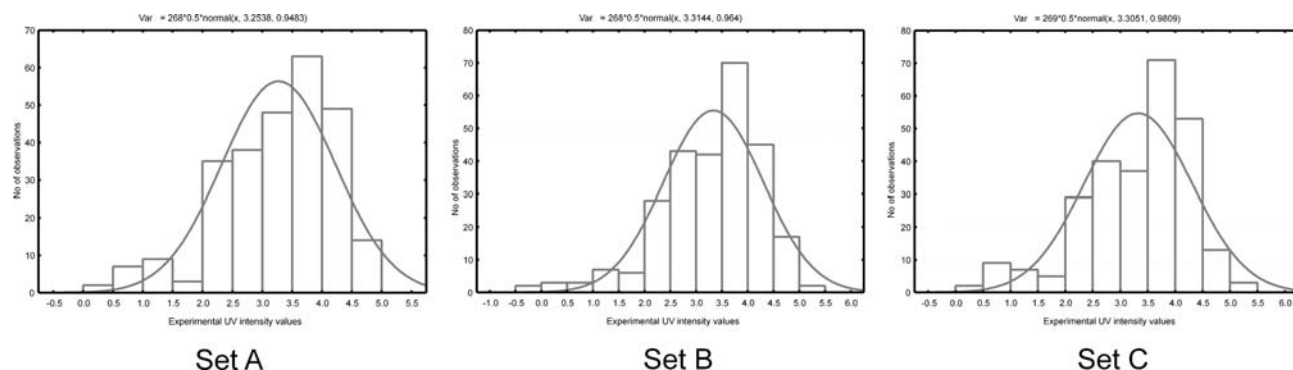

Fig. 2 The data distribution for the generated samples (subsets)

methodology section). The statistical parameters calculated by both approaches are given for the same subsets as they were defined.

BMLR results

Set A + B
$R^2 = 0.699$; $F = 245.960$; $s = 0.527$; $N = 5$; $n = 536$
$R^2 cvOO = 0.690$; $R^2 cvMO = 0.688$; RMS-PEOO = 0.530; RMSPEMO = 0.532
Ranges: Observed (0; –5.328) Predicted (–0.691; –5.1847)
External validation: set C; $n = 269$
$R^2 = 0.669$

Table 1 The derived BMLR model for dataset A + B

#	B	s	t	IC	Name of descriptor
0	3.637	0.39	9.329		Intercept
1	20.591	1.375	14.974	0.615	Square root of Partial Surface Area (MOPAC PC) for atom C
2	2.594	0.304	8.531	0.127	Relative number of double bonds
3	–0.245	0.034	–7.103	0.504	HOMO - LUMO energy gap
4	–5.553	0.888	–6.256	0.349	Moments of inertia C
5	0.029	0.005	5.724	0.231	count of H-donors sites (MOPAC PC)

ANN Results:
Set A + B
$R^2 = 0.764$; $S^2 = 0.217$;
External validation: set C
Coefficient of determination $R^2 = 0.672$; $S^2 = 0.344$

BMLR results:
Set A+C
$R^2 = 0.699$; $F = 246.070$; $s = 0.532$; $N = 5$; $n = 537$
$R^2 cvOO = 0.691$; $R^2 cvMO = 0.690$; RMS-PEOO = 0.540; RMSPEMO = 0.537
Ranges: Observed (0.062; –5.337) Predicted (0.513; –5.254)
External validation: set B; $n = 268$
$R^2 = 0.674$

Table 2 The derived BMLR model for dataset A + C

#	B	s	t	IC	Name of descriptor
0	2.017	0.367	5.502		Intercept
1	23.054	1.284	17.951	0.526	Square root of Partial Surface Area (MOPAC PC) for atom C
2	0.736	0.105	7.005	0.217	Average Information content (order 0)
3	–0.196	0.028	–6.984	0.397	HOMO - LUMO energy gap
4	0.108	0.017	6.370	0.096	Number of double bonds
5	–4.479	0.867	–5.166	0.238	Moments of inertia C

ANN Results:
Set A + C
$R^2 = 0.763$; $S^2 = 0.221$;
External validation: set B
Coefficient of determination $R^2 = 0.691$; $S^2 = 0.298$

BMLR results:
Set B + C
$R^2 = 0.687$; $F = 233.490$; $s = 0.546$; $N = 5$; $n = 537$
$R^2 cvOO = 0.679$; $R^2 cvMO = 0.680$; RMS-PEOO = 0.550; RMSPEMO = 0.549
Ranges: Observed (0; –5.337) Predicted (–0.337; –5.584)
External validation: set A; $n = 268$
$R^2 = 0.697$

Table 3 The derived BMLR model for dataset B+C

#	B	s	t	IC	Name of descriptor
0	2.483	0.385	6.444		Intercept
1	20.392	1.253	16.280	0.503	Square root of Partial Surface Area (MOPAC PC) for atom C
2	3.371	0.312	10.804	0.083	Relative number of double bonds
3	0.050	0.009	5.904	0.292	Bonding Information content (order 1)
4	–0.175	0.032	–5.533	0.465	HOMO - LUMO energy gap
5	–3.941	0.995	–3.960	0.442	Moments of inertia C

ANN Results:
Set B + C
$R^2 = 0.760$; $S^2 = 0.227$;
External validation: set A
Coefficient of determination $R^2 = 0.685$; $S^2 = 0.583$

The descriptors appearing in Tables 1, 2 and 3 for the submodels of datasets A + B, A + C and B + C are quite similar, with small differences due to the procedure applied for the descriptor selection in the BMLR method. Depending on the data set, different (but physically similar and highly intercorrelated) descriptors may appear in the different models. Namely, only one of a pair or a set of highly intercorrelated descriptors is used in the further model development

As can be seen from the models shown above [Tables 1, 2 and 3], three of the most statistically significant descriptors are the same for each of the three models (AB, AC and BC): Square root of Partial Surface Area (MOPAC PC) for atom C; HOMO – LUMO energy gap and Moments of inertia C. The presence of double bonds in the structures is underlined by the two descriptors ''relative number of double bonds'' and ''number of double bonds'' with the same physico-chemical meaning but different definitions.

The final discussed descriptors of each subset comprise: ''average information content (order 0)'', ''bonding information content (order 1)'' and ''count of H-donors sites (MOPAC PC)'' (see Tables 2, 3 and 1, respectively). The first two of these descriptors follow the tendency characterizing the mathematical nature of the BMLR algorithm in which one structural property appears under two different descriptor names in the models shown in Tables 2 and 3. The ''count of H-donors sites (MOPAC PC)'' is not from the group of the information content related descriptors. However, it is the least significant descriptor in the model shown in Table 1 and such difference is acceptable.

The resulting ANN models obtained for the same subsets are, as is to be expected, somewhat better than those derived by using BMLR. However, the ANN models, due to their nonlinear nature provide little information about the direction of the influence of the regression coefficient signs (there are no analogs) as the linear models. Thus, while the ANN models are quite useful for the prediction of the UV absorption values, they are difficult to interpret the observed relationship between the molecular structure characteristics and the property of interest.

In the next stage of the modeling process, we built a general QSPR models based only on the descriptors appearing in the submodels previously obtained [see Tables 1–3]. This subset of descriptors was further treated by BMLR procedure for all 805 compounds. The calculated statistical parameters of the five parameter model obtained and its predictive power are shown in Table 4 and Fig. 3.
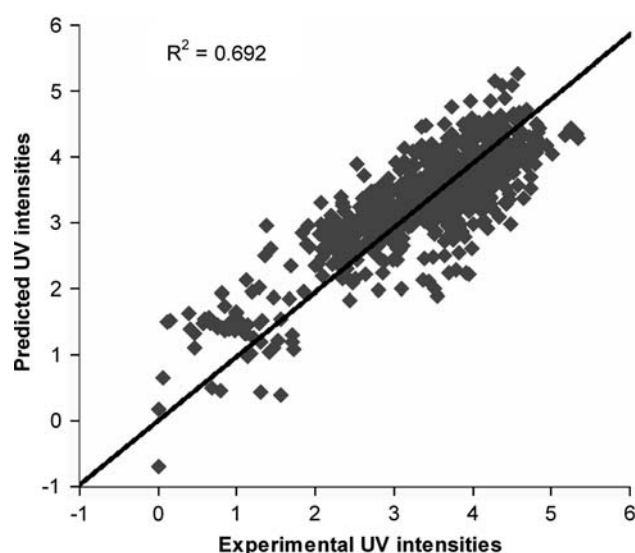

Fig. 3 BMLR predicted vs. experimental UV intensities values

Table 4 The derived BMLR model for whole dataset consisting of 805 molecules

#	B	s	t	IC	Name of descriptor
0	2.591	0.316	8.195		Intercept
1	20.507	1.024	20.017	0.513	Square root of Partial Surface Area (MOPAC PC) for atom C
2	2.442	0.258	9.470	0.151	Relative number of double bonds
3	–0.198	0.025	–7.823	0.448	HOMO - LUMO energy gap
4	–5.342	0.741	–7.206	0.334	Moments of inertia C
5	0.319	0.047	6.856	0.120	Average Information content (order 1)

$R^2 = 0.692$; $F = 358.830$; $s = 0.537$; $N = 5$; $n = 805$
$R^2_{cv}OO = 0.686$; $R^2_{cv}MO = 0.687$; RMSPEOO = 0.540; RMSPEMO = 0.540;
Ranges: Observed (0; –5.337) Predicted (–0.691; –5.262)

Table 5 Description of the statistical parameters used

Notation	Description
n	number of the compounds used
N	number of the descriptors used in the model
R^2	square of the correlation coefficient
s	standard deviation
F	Fisher criterion
B	linear regression coefficients
t	Student's criterion
IC	Partial intercorrelation coefficient
$R^2_{cv}OO$	square of the leave-one-out cross-validated correlation coefficient
$R^2_{cv}MO$	square of the leave-many-out cross-validated correlation coefficient
RMSPEOO	root mean square predictive error in the case of leave-one-out procedure used
RMSPEMO	root mean square predictive error in the case of leave-many-out procedure used

The derived model provides satisfactory statistical results (Table 4) considering the large number and diverse structure of the compounds. We selected the Table 4 equation from among the possible models for the following reason:

1. It provides a small number of easily and clearly interpretable descriptors which are physically reasonable in relation to the phenomenon studied. We were significantly guided during the selection of the model by our desire to obtain both the models with the best possible statistical parameters, and also those with understandable physico-chemical meaning.

2. The BMLR algorithm is such that by adding descriptors to the linear equation would lead to improvement of the statistical quality of the model. Having in mind that every additional descriptor in the model is less significant than those already included, we decided to stop at the point where the subsequent descriptors in the different subsets started to differ in physicochemical meaning. This was done in order to avoid the pure mathematical influence of the BMLR procedure over the modeling process (see the discussion above for the ''average information content'' and „count of H-donors sites'' descriptors).

3. The general model in Table 4, possesses almost identical R^2 and R^2_{cv}, a characteristic of stability. This is also true for the external validation datasets where R^2_{test} (R^2_{CV}MO) is close to R^2_{train}.

As an additional step, analysis of the model's domain of applicability for the general model [33, 34] was performed. A plot of the leverage vs. absolute residuals is shown in Fig. 4.
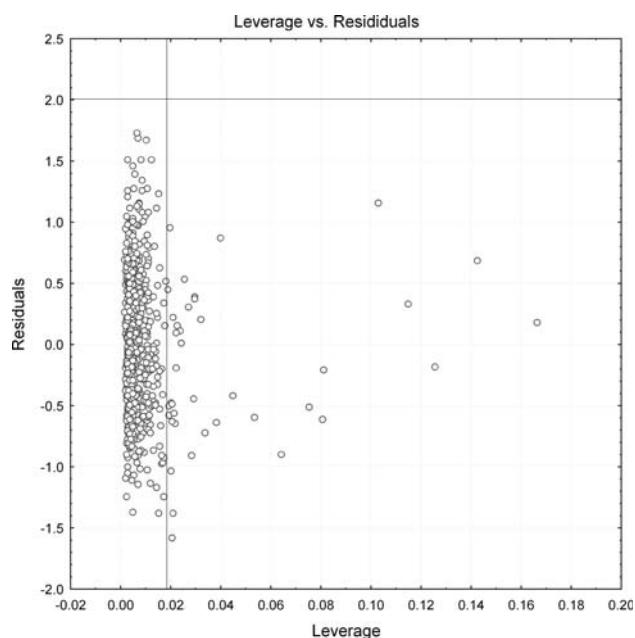

Fig. 4 Plot of a model's domain of applicability

As can be seen from Fig. 4 only about 4% of the structures (almost all containing sulfur atoms) show leverage greater than the threshold value ($3*N/n = 0.019$). A possible interpretation of this observation is the AM1 parameterization deficiency for the S atoms [35].

Each band in a UV spectrum can be represented by a combination of Gaussian and Lorenz analytical functions as in Eq. (1) [36]. Thus every single band can be represented by the following

$$f\,(v) = \mu\,f_G\,(v)\,+\,(1\,-\,\mu)\,f_L\,(v), \tag{1}$$

In Eq. (1), $f_G\,(v)$ and $f_L\,(v)$ are the Gaussian and Lorenz functions respectively, while μ is a variable, defined between 0 and 1 [36]. The application of these analytical functions is determined by their distinct physical significance and simple symmetrical shape. The use of asymmetrical functions would lead to an increase in the number of unknown parameters, and the limits of the asymmetry factor alteration cannot be determined.

Since each UV band could be represented by (1), the UV intensity can be calculated at any chosen wavelength λ, provided that the exact mathematical representation of this equation is known. From theoretical considerations, $f\,(v)$ depends on the following UV band parameters: intensity at the maximum (ε), the position (λ), and the half bandwidth ($v^{1/2}$). For any fixed λ close to the transition wavelength (having in mind that $f_G\,(v)$ and $f_L\,(v)$ are differentiable functions) ε is a unique function of these three parameters, i.e.

$$\varepsilon(260\,\mathrm{nm})\,=\,f\left(\varepsilon,\,\lambda,\,v_{1/2}\right) \tag{2}$$

Since, the function $f\,(v)$ is a good (but not exact) approximation of the band shape, one can expect that the relation between ε at certain λ based on Eq. (1) and (2) will include the approximation error. Thus, the correlation coefficient of the model proposed would be influenced by this approximation.

Searching for a robust, physically meaningful QSPR model, we tried to relate each of the descriptors that appeared in the model (Table 4), to the UV band parameters uniquely defining the function $f\,(v)$.

The ε at the transition wavelength λ depends largely on the following molecular properties (in terms of descriptors used): the relative number of the double bonds and the molecular shape and the symmetry. The bigger is the ''relative number of the double bonds'' descriptor, the bigger is the possibility for longer conjugated chains and higher extinction coefficient.

The ''Average Information content (order 1)'' descriptor describes the connectivity and molecule branching so, it is related to the molecular shape and symmetry. In general: the greater the molecular branching, the lower the symmetry. In highly symmetrical molecules, the bonding and anti-bonding orbitals possess almost the same symmetry (with respect to a corresponding element of symmetry) and can cause the transition to be of low intensity (''forbidden''). To summarize, it can be concluded that higher branching (positive regression coefficient in the model, see Table 4) defines lower symmetry and higher intensity.

The next variable in Eq. (2) is λ – the transition wavelength. It depends from the HOMO-LUMO energy gap

which defines the position of the transition maxima in the UV spectrum ($\Delta E = (v; \lambda = 1/v)$).

The final variable in Eq. (2) is $v_{1/2}$ - the half bandwidth (full width at half maximum). This is the most complicated parameter, since it is influenced by many parameters of both the compound and of the UV measuring apparatus. Thus, in place of rational estimate of the $v_{1/2}$ value, the simple empirical formula ($v_{1/2} = k\lambda$) is often used for its calculation. In this formula λ is the transition wavelength, and the parameter k is equal to 0.00375 [37]. This parameter together with the $f(v)$ function approximation causes the largest errors in the proposed model.

There are two more descriptors in the model presented in Table 4: ''Square root of Partial Surface Area (MOPAC PC) for atom C'' and ''Moments of inertia C''. The first one is likely related to the fact that the larger the number of C atoms (e.g. their surface area) the bigger the C-skeleton, which combined with the larger number of double bonds (both of them have positive coefficients) leads to the higher absorption.

The second descriptor is usually related to the mass distribution within the molecules. A possible interpretation for the ''Moments of inertia C'' descriptor in the model is its responsibility for the vibrational term of the excitation energy. The larger descriptor values correspond to the stronger inequality of the mass distribution in the molecules which lead to higher excitation energies and lower intensities (negative sign of the regression coefficient).

Thus, based on the above conjectures of the model proposed (with the described limitations for the $f(v)$ function and $v_{1/2}$) we believe this could be useful tool for the researchers using HPLC UV. Since, the mixtures of compounds identified by this method usually have $\Delta\varepsilon$ bigger than 1–2 log units, even this imperfect correlation ($R^2 = 0.692$) could be useful for identification purposes.

Conclusions

A modified QSPR approach combining the advantages of the two QSPR methods (linear and nonlinear) was applied to a set of 805 measured (260 nm) UV intensity values. Multilinear and nonlinear (ANN) QSPR equations with five theoretical molecular descriptors were obtained for the constructed subsets. A general multilinear model for all the compounds was proposed based on the results derived from the subsets. The descriptors involved are calculated solely from the chemical structures of the compounds and have definite physical meaning relative to the nature of the process.

References

1. Fitch WL, McGregor M, Katritzky AR, Lomaka A, Petrukhin R, Karelson M (2002) J Chem Inf Comp Sci 42:830
2. The MDL Drug Data Report (MDDR) is a Commercial Database Available from MDL Information Systems Inc., San Leandro
3. Williams GM, Carr RAE, Congreve MS, Kay C, McKeown SC, Murray PJ, Scicinski JJ, Watson SP (2000) Angew Chem Int Ed Engl 39:3293
4. Molnar SP, King JW (1997) Int J Quantum Chem 65:1047
5. Yanagita M, Kanda S, Tokita S (1999) Mol Cryst Liq Cryst Sci Technol Sect A 327:53
6. Türker L (2002) J Mol Struct (Theochem) 588:133
7. Horiguchi E, Shirai K, Matsuoka M, Matsui M (2002) Dyes Pigments 53:45
8. Al-Hazmy SM, Kassab KN, El-Daly SA, Ebeid EZM (2000) Spectrochim Acta A 56:1773
9. Machado AEH, Miranda JA, Guilardi S, Nicodem DE, Severino D (2003) Spectrochim Acta A 59:345
10. de Melo S, Fernandes PF (2001) J Mol Struct 565:69
11. Maud JM (1999) Synth Met 101:575
12. Breza M, Lukeš V, Vrábel I (2001) J Mol Struct (Theochem) 572:151
13. Lukeš V, Breza M, Laurinc V (2002) J Mol Struct (Theochem) 582:213
14. Lukeš V, Breza M, Végh D, Hrdlovič P, Krajčovič J, Laurinc V (2002) Synth Met 129:85
15. Ridley J, Zerner MC (1973) Theor Chim Acta 32:111
16. Internet reference. Retrieved from www.ccl.net/chemistry/resources/messages/1996/07/16.008-dir/index.html
17. Gross E, Dobson J, Petersilka M (1996) Top Curr Chem 181:81
18. Becke AD (1993) J Chem Phys 98:5648
19. CODESSA PRO Software, University of Florida, 2002
20. Karelson M, Maran U, Wang Y, Katritzky AR (1999) Collect Czech Chem Commun 64:1551
21. Katritzky AR, Taemm K, Kuanar M, Fara DC, Oliferenko A, Oliferenko P, Huddleston JG, Rogers RD (2004) J Chem Inf Comp Sci 44:136
22. Thakur A (2005) ARKIVOC 14:49
23. Basak SC, Mills D (2005) ARKIVOC 2:60
24. Internet reference. Retrieved from http://webbook.nist.gov/chemistry/
25. Hyperchem, v. 7.5; Hypercube Inc.; Gainesville, FL
26. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) J Am Chem Soc 107:3902
27. Katritzky AR, Ignatchenko ES, Barcock RA, Lobanov VS, Karelson M (1994) Anal Chem 11:1799
28. Karelson M (2000) Molecular Descriptors in QSAR/QSPR, Wiley-Interscience, New York
29. Zupan J, Gasteiger J (1999) Neural networks in chemistry and drug design, 2nd edn, Wiley-VCH, Weinheim
30. Rumelhart DE, Hinton GE, Williams RJ (1986) Parallel distributed processing: exploration in the microstructures of cognition. MIT Press, Cambridge
31. Svozil D, Kvasnicka V, Pospichal J (1997) J Chem Intel Lab Syst 39:43
32. Erikson L, Jaworska J, Worth AP, Cronin MTD, McDowell RM (2003) Environm Health Persp 111:1361
33. Gonzalez-Diaz H, Vilar S, Santana L, Podda G, Uriarte E (2007) Bioorg Med Chem 15(7):2544
34. Winget P, Horn AHC, Selcuki C, Martin B, Clark T (2003) J Mol Model 9:408
35. Antonov L, Stoyanov S (1993) Appl Spectr 47:1030
36. Voloshina EN, Raabe G, Estermeier M, Steffan B, Fleischhauer J (2004) Int J Quantum Chem 100:1104