J Mol Model (2005) 12: 65–75
DOI 10.1007/s00894-005-0006-x

ORIGINAL PAPER

Jie Xu · Bin Guo · Biao Chen · Qijin Zhang

A QSPR treatment for the thermal stabilities of second-order NLO chromophore molecules

Received: 10 January 2005 / Accepted: 27 June 2005 / Published online: 21 October 2005
© Springer-Verlag 2005

Abstract Quantitative structure–property relationships were studied between descriptors representing the molecular structures and thermal decomposition temperatures (T_d) for a diverse set of 90 second-order nonlinear optical (NLO) chromophores. A seven-parameter model was developed for the prediction of molar thermal decomposition function Y_d (T_d M, where M represents the molar weight) with R^2 = 0.9642 and SEE = 14.01 by multilinear regression analysis. The mean relative error for the prediction of T_d was 4.46%. The stability of the proposed model was validated using leave-one-out cross-validation. All descriptors involved in the model were derived solely from the chemical structures of the NLO chromophores.

Keywords QSPR · Thermal stabilities · NLO chromophores · Molecular descriptors · Multilinear regression analysis

Abbreviations QSPR: Quantitative structure–property relationships · NLO: Nonlinear optical · SEE: Standard error of estimation · LOO: Leave-one-out · MLRA: Multilinear regression analysis

Introduction

The increasing importance of telecommunications has caused a pressing need to develop photonic components, which are actually fabricated from semiconductors. The performances of semiconductor devices are not optimal and the interest in organic second-order nonlinear optical (NLO) materials is growing because of their larger susceptibilities, faster response time, ease of processing, and versatility of molecular structural modifications. In particular, poled polymeric materials will 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 small organic molecules, namely chromophores, incorporated into polymer matrices 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—$\mu\beta$, where μ is the dipole moment and β is the hyperpolarizability); (2) good solubility of the chromophore in polymer matrices to avoid phase separation and formation of concentration gradients; (3) good thermal stability, usually with the thermal decomposition temperature T_d over 573 K to ensure that the chromophores are sufficiently thermally robust to survive the high temperatures (523–573 K) during the fabrication steps for those devices; (4) low cut-off wavelength and so on [4–7]. Optimization of any individual above-mentioned property is not difficult. However, simultaneous realization of all requirements to achieve device-quality materials is not a trivial task [6]. Due to the existing trade-offs between these properties, optimization of one very often causes the attenuation of others. For example, additional thermal stability can be added to a NLO chromophore by the substitution of aromatic moieties for aliphatic ones along the delocalized path between an electron donor and an acceptor. However, since the molecular nonlinearity is a measure of the ease of excitation to a state in which substantial intramolecular charge transfer has taken place, additional aromatic rings disfavor that charge transfer and the molecular nonlinearity is reduced [8]. Therefore, the design and synthesis of optimized chromophores have

J. Xu
Structure Research Laboratory,
University of Science and Technology of China,
Hefei, 230026, P. R. China

B. Guo · B. Chen · Q. Zhang (✉)
Department of Polymer Science and Engineering,
University of Science and Technology of China,
Hefei, 230026, P. R. China

become a key goal in this area of research [8–12]. In addition, methods for predicting the above-mentioned properties of chromophores from their molecular structures quantitatively would be undoubtedly of significant utility in the search for device-quality chromophores and materials.

Alternatively, quantitative structure–property relationships (QSPR) provide a promising method for the estimation of the properties 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 the molecular features of the compounds termed descriptors. The advantage of this approach lies in the fact that it requires only knowledge of the chemical structure and is not dependent on any experimental properties. QSPR has been applied successfully to the correlation of many diverse physicochemical properties of chemical compounds. Remarkably, Oberg et al. [13] developed a successful QSPR model for the prediction of the nonlinearities of 22 chromophores, with 5 quantum-chemical descriptors involved. More recently, Zeng et al. [14] obtained a three-parameter correlation to predict the nonlinearities for 32 *para*-disubstituted benzenes. However, there have been relatively few attempts to predict T_d of compounds. The only previously published QSPR relationship for the prediction of T_d was developed by Bicerano [15] for a set of 140 polymers, with 21 descriptors involved, using the molar thermal decomposition function Y_d (defined as T_d multiplied by the molecular weight M) as the dependent variable ($R^2 = 0.998$).

The goal of the present study is (1) to develop a robust QSPR model that could predict T_d values for a diverse set of chromophores, (2) to discover the main structural factors that affect the T_d of chromophores significantly. The molecular structures of 90 chromophores were preoptimized and 444 descriptors were calculated for each molecule using the Dragon software [16]. Through stepwise multilinear regression analysis (MLRA) [17] with leave-one-out (LOO) cross-validation, a seven-parameter correlation was produced for Y_d, to allow the estimation of this quantity, and hence also of the T_d values of chromophores.

Materials and methods

Database construction

The molecular structures of the 90 NLO chromophore molecules with extensive structural diversity are shown in Fig. 1. The experimental T_d values were taken from several publications [7, 9–12, 18–29]. To select significant descriptors for the QSPR model that captures all the underlying interaction mechanisms, it is advisable to have as many structural features as possible represented in the data set. The working data set included traditional D-π-A, octupolar, Λ -shaped, Y-shaped, X-shaped, 2D charge-transfer chromophores, etc. Also, the compounds chosen contained various functionalities, such as aldehyde, amino, azo, cyano, ether, ethynyl, furan, nitro, phenyl, pyrrole, thiazole, thiophene, and so on. The structures included C, H, N, O, S, P, and F atoms. The molecular weights of these compounds ranged from 266.29 to 744.47 amu. The reported T_d values ranged from 473 to 685 K (see Table 1). Since Bicerano [15] obtained successful results in the prediction of T_d of polymers by using Y_d as the dependent variable because T_d is itself an intensive property, while Y_d is an extensive property, this approach was introduced.

Structure entry and descriptor generation

The structures of the molecules were sketched on a PC using the ChemDraw Ultra 7.0 program in ChemOffice 2002 [30]. The three-dimensional (3D) geometries were further preoptimized with the semiempirical AM1 method in the Chem3D program to ensure that low-energy conformations were obtained for each structure. The resulting geometry of each structure was transferred into the software Dragon [16] to calculate 444 empirical descriptors. The molecular descriptors thus generated include topological descriptors, molecular walk 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].

Objective feature selection

Objective feature selection was used 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 those that contain identical information for over 90% of the molecules were removed. Pairwise correlation analysis of the remaining descriptors using 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 0.90, regressions were built using descriptor subsets containing only one of these highly correlated descriptors. The descriptor with lower t value was removed. In this step, 313 descriptors were removed, resulting in the working subset with 131 descriptors.

Model development and validation

To develop QSPR models, stepwise MLRA [17] was applied to the data set. Stepwise MLRA produces a multiple-term linear equation. However, not all inde-

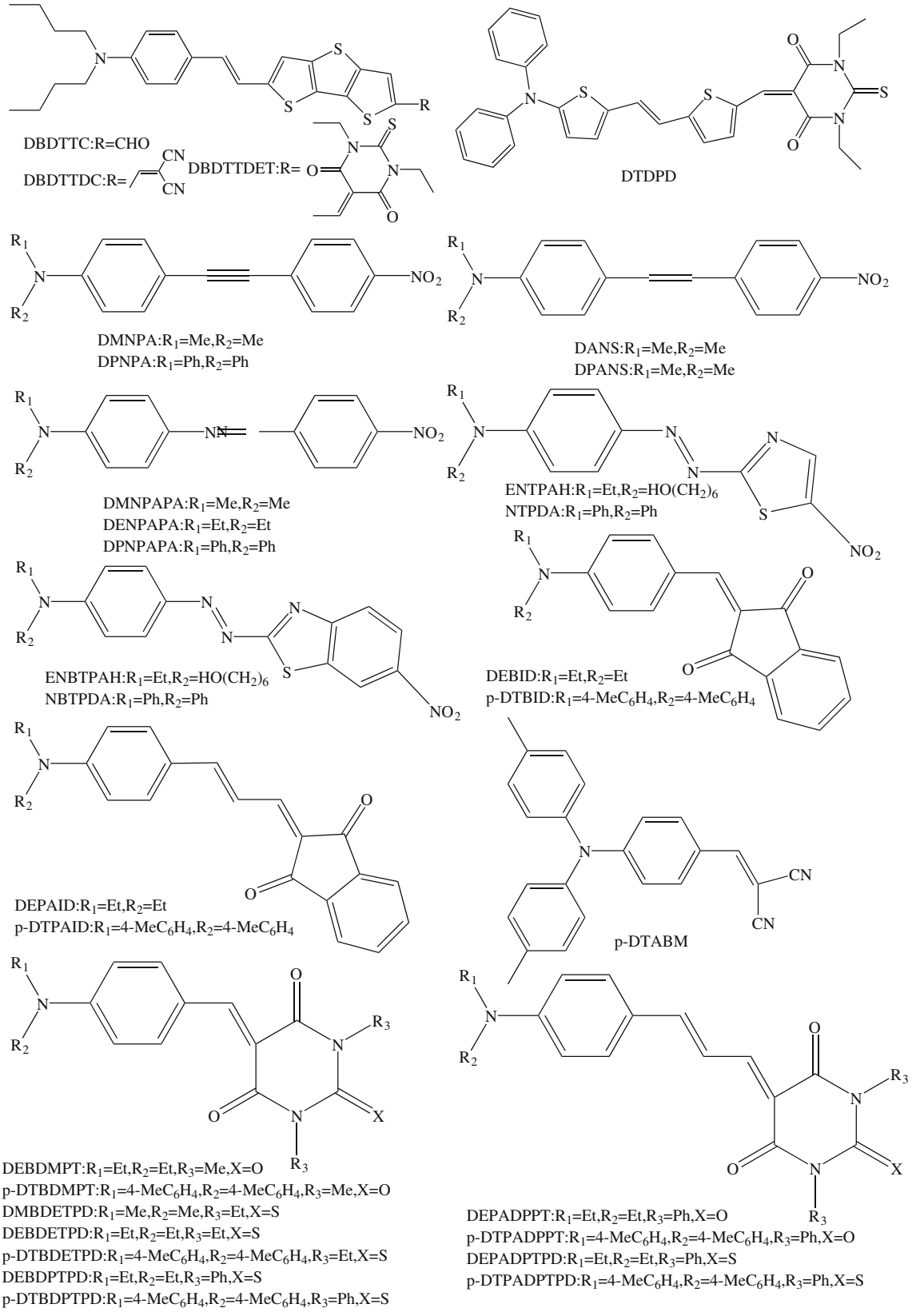


Fig. 1 The chromophores' structures included in the data set

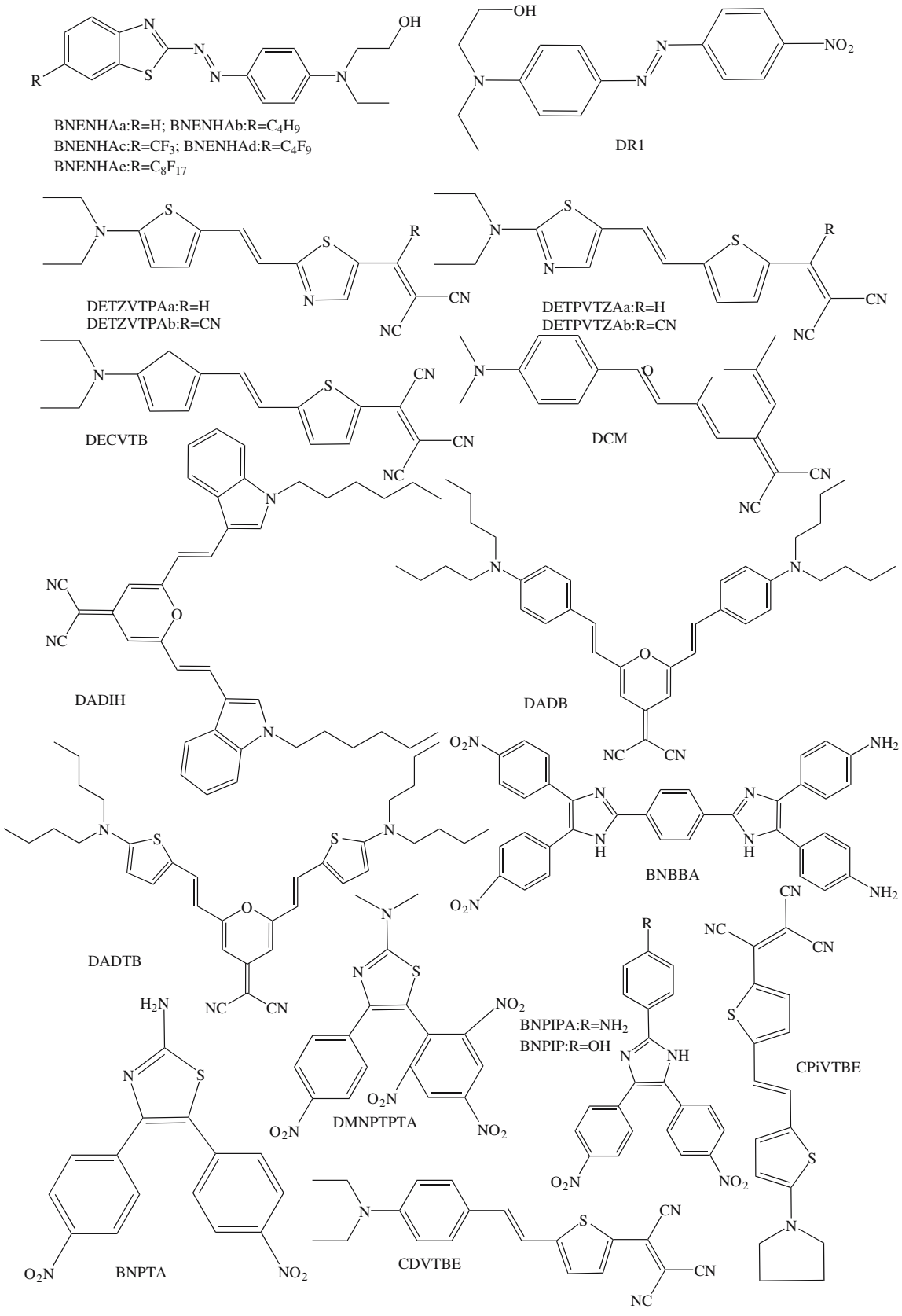

BNENHAa:R=H; BNENHAb:R=C4H9
BNENHAc:R=CF3; BNENHAd:R=C4F9
BNENHAe:R=C8F17
DR1
DETZVTPAa:R=H
DETZVTPAb:R=CN
DETPVTZAa:R=H
DETPVTZAb:R=CN
DECVTB
DCM
DADIH
DADB
DADTB
BNBBA
BNPIPA:R=NH2
BNPIP:R=OH
CPiVTBE
DMNPTPTA
BNPTA
CDVTBE

Fig. 1 (Contd.)

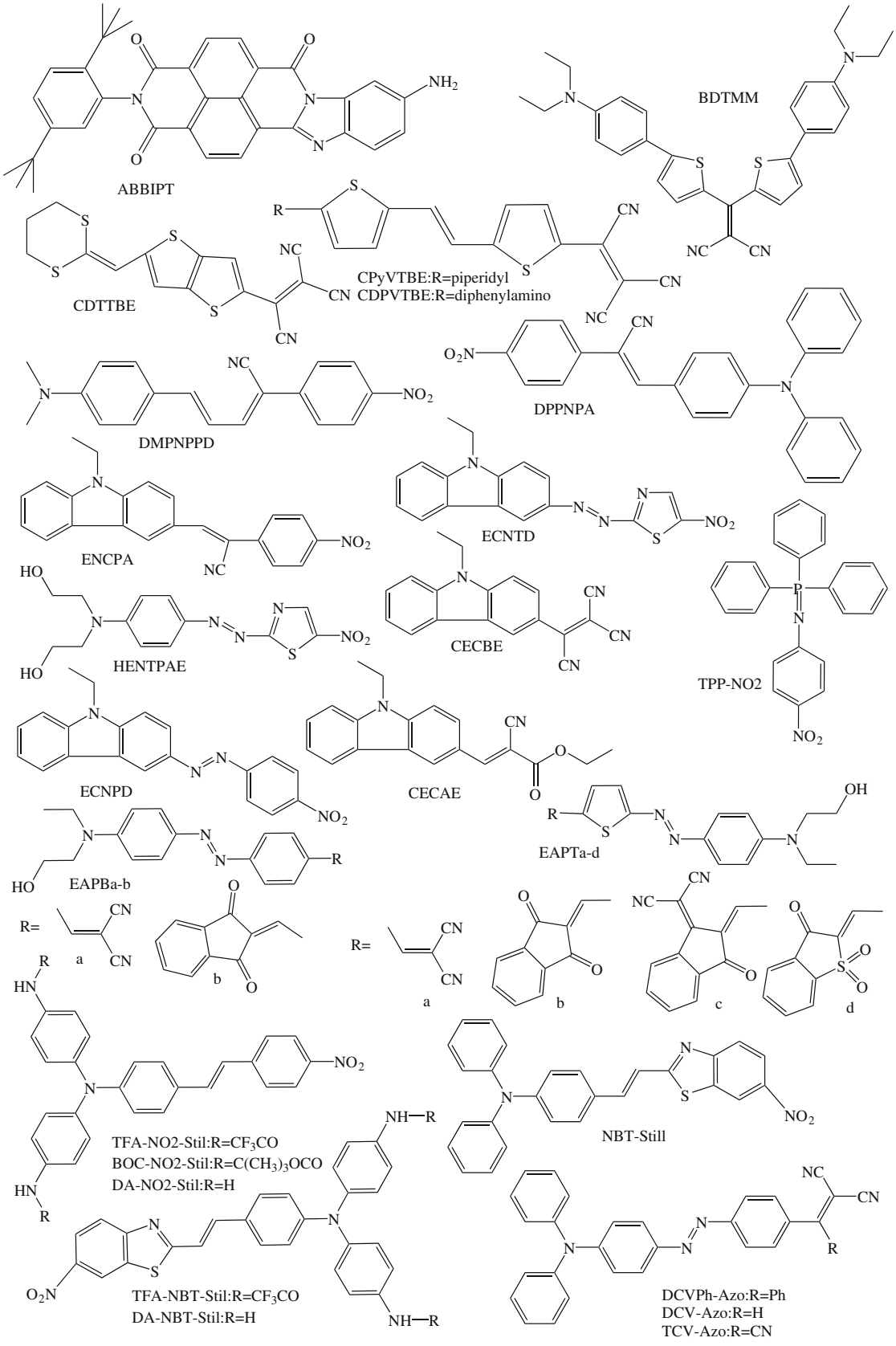
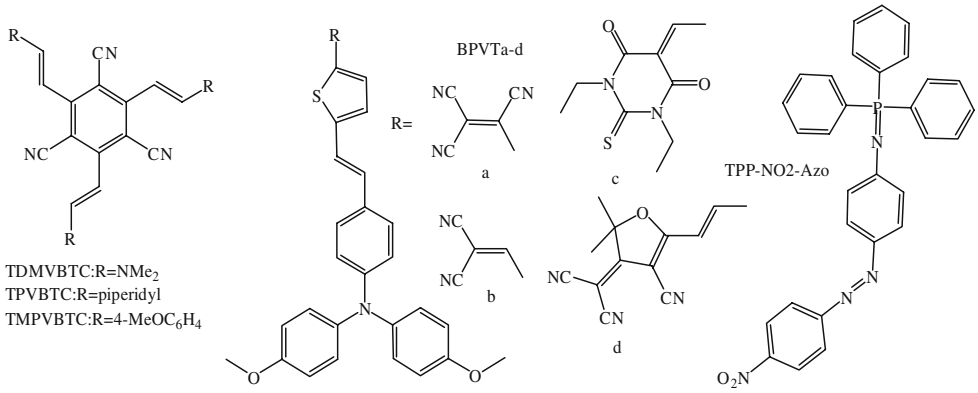

ABBIPT
BDTMM
CDTTBE
CPyVTBE:R=piperidyl
CDPVTBE:R=diphenylamino
DMPNPPD
DPPNPA
ENCPA
ECNTD
HENTPAE
CECBE
TPP-NO2
ECNPD
CECAE
EAPTa-d
EAPBa-b
TFA-NO2-Stil:R=CF3CO
BOC-NO2-Stil:R=C(CH3)3OCO
DA-NO2-Stil:R=H
NBT-Still
TFA-NBT-Stil:R=CF3CO
DA-NBT-Stil:R=H
DCVPh-Azo:R=Ph
DCV-Azo:R=H
TCV-Azo:R=CN

Fig. 1 (Contd.)

Fig. 1 (Contd.)



pendent variables are used. 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 overfitting. *F*-to-enter and *F*-to-remove were 4 and 3, respectively. The QSPR regression equation was tested by the coefficient regression (R^2), the adjusted R^2, the *F* ratio values, the standard error of estimates (SEE), and *p* values all corresponding to a 95% confidence level. The adjusted R^2 value was calculated using the following formula:

$$R^2_{\text{adj}} = 1 - \left[\left(\frac{n-1}{n-m-1}\right)\left(1 - \frac{\sum_i^n \left(Y_{\text{d},i}^{\text{calc}} - \overline{Y}_{\text{d}}\right)^2}{\sum_i^n \left(Y_{\text{d},i}^{\text{exp}} - \overline{Y}_{\text{d}}\right)^2}\right)\right], \tag{1}$$

where *n* is the number of objects and *m* the number of descriptors involved in the correlation while,

$$R^2 = \frac{\sum_i^n \left(Y_{\text{d},i}^{\text{calc}} - \overline{Y}_{\text{d}}\right)^2}{\sum_i^n \left(Y_{\text{d},i}^{\text{exp}} - \overline{Y}_{\text{d}}\right)^2}. \tag{2}$$

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., $R^2 = 1$ [33]. 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:

$$\text{MRE} = \frac{100}{n}\sum_i \left|\frac{T_{\text{d},i}^{\text{exp}} - T_{\text{d},i}^{\text{calc}}}{T_{\text{d},i}^{\text{exp}}}\right|. \tag{3}$$

The stability of the final QSPR model was further validated internally using LOO cross-validation. To do this, one compound of the data set was removed, and the model was recalculated using the remaining *n*-1 compounds as a training set. The property was then predicted for the excluded element. This process was repeated for all the compounds of the data set, obtaining a prediction for every one. From the residual values obtained, the standard error of estimates for the cross-validation SEE(CV) was determined.

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 the aforementioned descriptors 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 Y_{d}.

Results and discussion

The number of descriptors in the final QSPR model was determined on the basis of the data set size (90 chromophores) and on the basis of the correlation coefficient obtained, *F* ratio values, and adjusted correlation coefficient. In this way, the best seven-parameter correlation obtained for the entire data set of 90 chromophores had a squared correlation coefficient $R^2 = 0.9642$ and an adjusted correlation coefficient $R^2_{\text{adj}} = 0.9611$, as shown in detail in Table1 and Fig. 2. The MLRA representation of the Y_{d} is the following:

$$\begin{aligned} Y_d = {} & 0.869\text{VRD1} - 40.688\text{GATS4v} + 55.969\text{MATS6m} \\ & + 6.222\text{X5sol} - 39.544\text{EEig06r} + 2137.1\text{JGI7} \\ & - 1340.1\text{JGI8} + 157.722 \end{aligned} \tag{4}$$

Here VRD1 is the Randic-type eigenvector-based index from the distance matrix (D) [34]; GATS4v is the Geary autocorrelation –lag 4/weighted by atomic van der Waals volumes [31]; MATS6m is the Moran

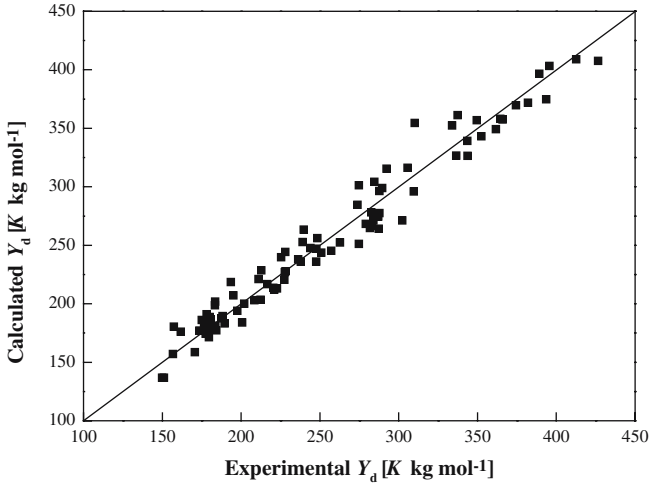


Fig. 2 Plot of predicted versus experimental values of Y_d obtained with seven descriptors involved ($R = 0.982$)

autocorrelation –lag 6/weighted by atomic masses [31]; X5sol is the solvation connectivity index chi-5 [35]; EEig06r is the eigenvalue 06 from the edge-adjacency matrix weighted by resonance integrals [36–40]; and JGI7 and JGI8 are the mean topological charge indices of orders 7 and 8, respectively [41].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 $R^2 = 0.9642$ indicates the quality of the selected model. The SEE results during stepwise MLRA were shown in Fig. 3, which indicate that equations with fewer descriptors showed poorer SEEs. For example, the simplest equation with $R^2 > 0.9$ contained only one descriptor (VRD1). This equation showed $R^2 = 0.9210$, but SEE = 20.08, i.e., a standard error of the estimates 43% greater than the one obtained with the best seven-parameter equation. Models containing more than seven descriptors did not show significantly better results than the seven-descriptor model.To test the dependency of the Y_d on each descriptor, R^2, R^2_{adj} and SEE were calculated with one-descriptor linear regression for the whole set of 90 chromophores (see Table 3). The first important descriptor is VRD1, which gave the best overall fit with an R^2 value of 0.9210 for the whole 90 data set. This descriptor encodes information about the size and the intramolecular ordering of the molecule. The positive sign of the VRD1 indicates that the bigger the size and the higher the ordering of the NLO chromophores, the better the thermal stability.The second important descriptor is the solvation connectivity index chi-5 (X5sol), which could be considered as entropy of solvation and somehow indicates the dispersion interactions occurring in the melt phase. The contributions of this descriptor in the thermal stability are in agreement with the contributions that one would expect for the interactions of the chromophores. X5sol is also a measure of branching of the molecules. Thus the presence of X5sol in the equation also indicates that the thermal stability depends on the degree of branching and compactness of the chromophore molecules. The index increases with increased chain branching. The coefficient for the X5sol descriptor is positive, meaning that as branching increases, the T_d increases.The presence of the descriptor EEig06r in the equation illustrates the influence of aromatic functionalities and their resonances on thermal stability. This descriptor decreases with increased resonances. The negative sign of EEig06r indicates that the resonance structures would enhance the T_d of the chromophores. The importance of the transfers of intramolecular charge on the thermal stability is apparent from the presence of JGI7 and JGI8 in the equation.Leave-one-out cross-validation was used to test the stability of the model obtained and the results are shown in Table 1 (columns 5 and 8) and Fig. 4. The discrepancies between both relative errors (RE) are small for most of the chromophores studied. This indicates that the model is reliable for the prediction of the T_d of NLO chromophores. The mean errors obtained by Eq. 4 and for the cross-validation were 25.2 and 27.6 K. The MREs were 4.46 and 4.89%, respectively. These values are acceptable if the uncertainty that accompanies the experimental determination of T_d for each case is taken into consideration. Usually the T_d values for chromophores are measured by thermogravimetric analysis (TGA) or differential scanning calorimetry (DSC). However, the discrepancies between reported results in the literature can be quite large. Take TPP-NO_2 for example [27], the T_d values determined by TGA or DSC are 555 or 600 K, respectively. In addition, the T_d values obtained are strongly affected by the heating rate and other conditions in the TGA or DSC measurement.The same model size and algorithm that produced the best model for the standard experiment were tested with randomized dependent variables. The most predictive model with SEE of 64.12 ($R^2 = 0.250$, $R^2_{adj} = 0.186$) was obtained with random dependent variables. The SEE and R^2 values indicate that a poor correlation was found between structure and Y_d, which proves the validity of the real model.From Bicerano's model to predict T_d of polymers [15], the two most significant descriptors, the number of non-hydrogen atoms and the first-order valence connectivity index $^1\chi^v$ of the repeat unit encode information about the size and branching of the repeat unit, which are expressed by the VRD1 and X5sol in our model. Sometimes the glass transition temperature T_g is also used as a measure of the thermal stability [42, 43] because there are many similarities between the trends observed for T_g and for T_d [15]. Kim et al. [42] obtained a seven-parameter QSPR model with $R^2 = 0.989$ for T_gs of small molecules by genetic algorithm (GA) and multilinear regression. The most significant factor, the number of bonds in a molecule (SC-1), is related to the size of the molecule, which is coded by VRD1. More recently, Yin et al. [43] studied the thermal stabilities of organic light emitting diode (OLED) materials using T_g as the dependent

Table 1 Prediction of the molar thermal decomposition function Y_d (Kkgmol^{-1}) and the decomposition temperature T_d (K) of 90 NLO chromophores

Chromophore	M	Y_d (exp)	Y_d (cal)[a]	Y_d (CV)[b]	T_d (exp)	T_d (cal)[c]	T_d (CV)[d]
DBDTTC[11]	453.68	281.80	264.68	262.62	621	583.4	578.9
DBDTTDC[11]	501.73	309.64	295.89	294.35	617	589.7	586.7
DBDTTDET[11]	635.93	333.96	352.30	356.10	525	554.0	560.0
DTDPD[19]	569.76	289.52	298.91	299.47	508	524.6	525.7
DMNPA[9]	266.29	149.96	136.71	135.15	563	513.4	507.6
DPNPA[9]	390.43	237.83	235.99	235.82	609	604.4	604.0
DANS[9]	268.31	151.10	136.71	134.90	563	509.6	502.8
DPANS[9]	392.45	247.69	235.99	235.51	631	601.4	600.1
DMNPAPA[9]	270.29	156.81	156.99	157.02	580	580.9	581.0
DENPAPA[9]	298.34	177.56	179.72	180.02	595	602.5	603.4
DPNPAPA[9]	394.43	262.75	252.29	251.27	666	639.7	637.1
ENTPAH[9]	377.46	183.50	198.62	199.16	486	526.2	527.7
NTPDA[9]	401.44	228.08	244.26	245.20	568	608.5	610.8
ENBTPAH[9]	427.52	227.50	220.52	220.14	532	515.9	515.0
NBTPDA[9]	451.50	284.06	271.21	270.40	629	600.7	598.9
DEBID[9]	305.37	170.75	158.53	157.70	559	519.2	516.5
p-DTBID[9]	429.51	287.41	264.19	263.37	669	615.1	613.2
DEPAID[9]	331.41	180.01	188.59	188.95	543	569.1	570.2
p-DTPAID[9]	455.55	273.85	284.52	285.16	601	624.6	626.0
p-DTABM[9]	349.43	228.23	227.82	227.77	653	652.0	651.9
DEBDMPT[9]	315.37	157.42	180.47	183.73	499	572.3	582.6
p-DTBDMPT[9]	439.51	282.67	278.07	277.46	643	632.7	631.3
DMBDETPD[9]	331.43	161.79	176.07	177.37	488	531.3	535.2
DEBDETPD[9]	359.49	176.56	183.33	183.50	491	510.0	510.5
p-DTBDETPD[9]	483.62	302.34	271.15	269.45	625	560.7	557.2
DEBDPTPD[9]	455.17	244.04	247.65	247.72	536	544.1	544.3
p-DTBDPTPD[9]	579.71	349.65	356.76	357.44	603	615.5	616.6
DEPADPPT[9]	463.57	250.86	243.67	242.90	541	525.7	524.0
p-DTPADPPT[9]	587.71	337.43	361.08	363.09	574	614.4	617.9
DEPADPTPD[9]	479.63	257.15	245.16	243.80	536	511.2	508.4
p-DTPADPTPD[9]	603.77	365.98	357.44	356.01	606	592.1	589.7
BNENHAa[21]	326.42	179.58	171.36	170.61	550	525.0	522.7
BNENHAb[21]	382.52	212.36	203.40	202.88	555	531.8	530.4
BNENHAc[21]	394.41	202.00	199.99	199.76	512	507.1	506.5
BNENHAd[21]	544.44	283.19	266.54	264.20	520	489.6	485.3
BNENHAe[21]	744.47	382.02	371.83	367.60	513	499.5	493.8
DR1[21]	314.34	179.85	178.24	178.04	572	567.1	566.5
DETZVTPAa[20]	365.08	178.94	185.16	185.43	490	507.2	508.0
DETZVTPAb[20]	340.47	173.69	176.98	177.39	510	519.8	521.1
DETPVTZAa[20]	365.48	188.28	189.47	189.49	515	518.5	518.5
DETPVTZAb[20]	340.47	184.25	177.16	176.28	541	520.4	517.8
DECVTB[20]	346.45	179.51	186.63	186.85	518	538.7	539.4
DCM[10]	303.36	183.58	201.82	203.87	605	665.3	672.1
DADIH[10]	594.79	395.62	403.22	405.36	665	678.0	681.6
DADB[10]	602.85	374.46	369.61	368.97	621	613.2	612.1
DADTB[10]	614.91	352.44	343.01	342.15	573	557.9	556.5
BNBBA[23]	634.64	412.61	408.88	408.06	650	644.3	643.0
BNPTA[25]	342.33	200.66	184.00	181.35	586	537.6	529.8
DMNPTPTA[25]	460.38	236.24	237.99	238.13	513	517.0	517.3
BNPIPA[25]	401.37	222.82	213.03	211.49	555	530.8	527.0
BNPIP[25]	402.36	193.60	218.57	221.95	481	543.3	551.7
CDVTBE[22]	358.46	187.53	187.38	187.38	523	522.8	522.8
CPiVTBE[22]	362.47	189.63	183.08	182.67	523	505.1	504.0
ABBIPT[22]	542.63	343.57	339.08	337.62	633	624.9	622.2
BDTMM[22]	536.75	336.62	326.64	326.63	627	608.6	608.6
CDTTBE[22]	371.52	216.65	216.73	216.74	583	583.4	583.4
CPyVTBE[18]	376.50	178.14	190.90	191.38	473	507.1	508.4
CDPVTBE[18]	460.57	247.86	247.08	246.98	538	536.5	536.3
DMPNPPD[28]	319.36	183.68	181.21	180.96	575	567.5	566.7
DPPNPA[28]	417.46	274.75	251.10	250.44	658	601.5	600.0
ECNPA[28]	367.40	227.84	226.88	226.74	620	617.6	617.2
ECNTD[7]	351.38	208.42	202.98	202.43	593	577.7	576.1
HENTPAE[7]	337.35	175.14	186.12	187.36	519	551.7	555.4
CECBE[7]	296.33	197.70	194.02	193.49	667	654.8	653.0
ECNPD[7]	344.37	220.45	213.50	212.61	640	620.0	617.4
CECAE[7]	318.37	195.21	207.02	208.27	613	650.3	654.2

Table 1 (Contd.)

Chromophore	M	Y_d (exp)	Y_d (cal)[a]	Y_d (CV)[b]	T_d (exp)	T_d (cal)[c]	T_d (CV)[d]
EAPBa[26]	345.40	180.87	186.74	186.87	524	540.7	541.1
EAPBb[26]	425.48	212.80	228.53	229.15	500	537.2	538.6
EAPTa[26]	351.43	177.52	174.33	174.13	505	496.1	495.5
EAPTb[26]	431.51	211.07	221.06	221.56	489	512.3	513.5
EAPTc[26]	479.55	239.85	263.25	263.98	500	549.0	550.5
EAPTd[26]	467.56	225.43	239.66	240.57	482	512.6	514.6
TFA-NO2-Stil[24]	614.49	364.48	357.95	356.75	593	582.6	580.6
BOC-NO2-Stil[24]	622.71	310.20	354.40	360.86	498	569.2	579.5
DA-NO2-Stil[24]	422.48	248.48	256.08	256.51	588	606.2	607.2
TFA-NBT-Stil[24]	671.57	426.55	407.52	405.02	635	606.8	603.1
DA-NBT-Stil[24]	479.55	287.80	296.33	296.70	600	618.0	618.8
NBT-Stil[24]	449.52	287.76	277.51	276.84	640	617.4	615.9
DCVPh-Azo[24]	501.58	343.66	326.49	325.73	685	651.0	649.5
DCV-Azo[24]	425.48	279.18	268.30	267.71	656	630.6	629.2
TCV-Azo[24]	450.49	287.03	274.43	274.03	637	609.2	608.3
TDMVBTC[12]	360.46	221.02	212.07	210.18	613	588.4	583.1
TPVBTC[12]	480.65	284.62	304.11	308.68	592	632.8	642.3
TMPVBTC[12]	549.62	361.73	349.09	347.87	658	635.2	633.0
BPVTa[29]	514.60	292.37	315.36	316.31	568	612.9	614.7
BPVTb[29]	489.59	274.73	301.23	303.73	561	615.3	620.4
BPVTc[29]	623.78	393.70	374.73	373.17	631	600.8	598.3
BPVTd[29]	622.73	389.30	396.587	397.17	625	636.9	637.8
TPP-NO2[27]	398.39	239.09	252.657	254.55	600	634.2	639.0
TPP-NO2-Azo[27]	502.50	305.60	316.17	317.06	608	629.2	631.0

[a] Predicted using Eq. 4 (see text)
[b] Calculated from LOO cross-validation
[c] Calculated by T_d (cal) = Y_d (cal)/M
[d] Calculated by T_d (cal) = Y_d (cal)/M

Table 2 Descriptors involved in the best seven-parameter correlation derived for Y_d

Descriptor	Descriptor type	X	DX	*t* test	*p* level
Constant		157.722	20.009	7.882	0.000000
VRD1	Eigenvalue-based	0.869	0.049	17.736	0.000000
GATS4v	2-D autocorrelations	−40.688	7.897	−5.153	0.000002
MATS6m	2-D autocorrelations	55.969	8.621	6.492	0.000000
X5sol	Connectivity	6.222	1.237	5.031	0.000003
EEig06r	Edge-adjacency	−39.544	8.485	−4.660	0.000012
JGI7	Topological charge	2137.1	614.00	3.481	0.000804
JGI8	Topological charge	−1340.1	465.82	−2.877	0.005119

$n = 90$, $R^2 = 0.9642$, $R^2_{adj} = 0.9611$, $F = 315.25$, SEE = 14.01, SEE(CV) = 14.89, $p < 0.00000$

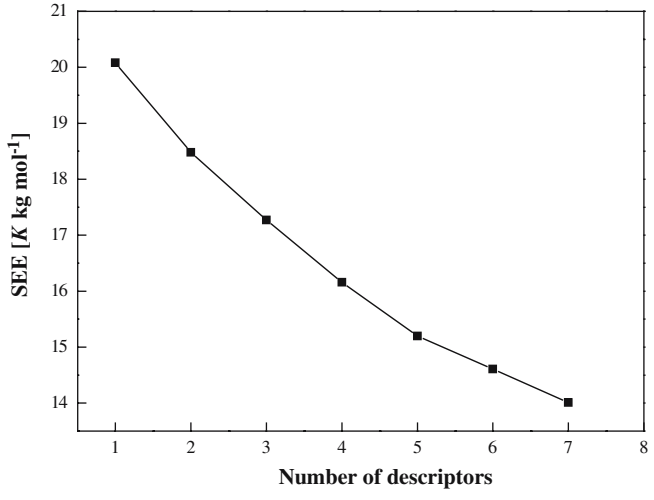


Fig. 3 SEE versus number of descriptors in the best MLRA equation for the whole set of chromophores

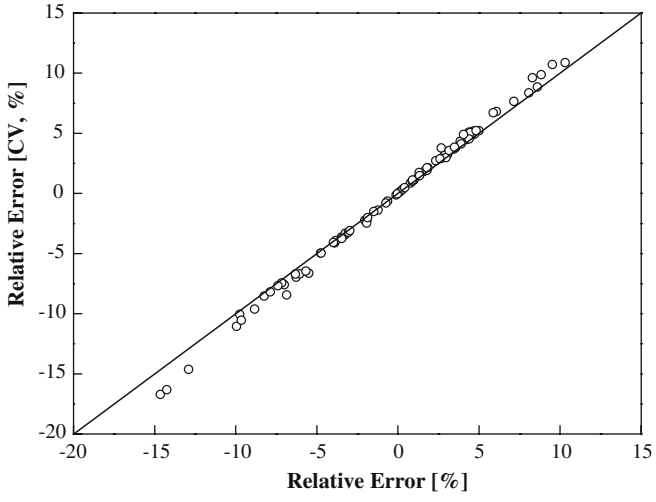


Fig. 4 Relative error obtained with the best MLRA versus RE obtained by cross-validation for T_d

Table 3 Summary of one descriptor models of 90 data set

Descriptor	Y_d prediction equation	R^2	R^2_{adj}	SEE (Kkgmol^{-1})
VRD1	$Y_d = 69.362 + 0.859$VRD1	0.921	0.920	20.08
X5sol	$Y_d = 59.267 + 24.683$X5sol	0.731	0.728	37.04
EEig06r	$Y_d = -160.520 + 140.797$EEig06r	0.718	0.715	37.95
JGI8	$Y_d = 325.185 - 7225.335$JGI8	0.135	0.125	66.47
MATS6m	$Y_d = 252.039 + 30.789$MATS6m	0.006	−0.005	71.24
JGI7	$Y_d = 228.946 + 1756.332$JGI7	0.005	−0.007	71.30
GATS4v	$Y_d = 271.211 - 20.428$GATS4v	0.004	−0.008	71.34

variable by means of QSPR. Using the CODESSA (comprehensive descriptors for structural and statistical analysis) package, they produced a six-parameter correlation with a mean error of 8.5 K for 73 OLED materials. The relative number of rings (RNR) and the principal moment of the inertia I_B in their model have similar physical meanings as EEig06r and VRD1. The similarities mentioned above indicate that the main structural factors affecting the thermal stability have been expressed by our model.

Conclusions

In this paper, a general seven-parameter QSPR model was reported to predict the T_d values for a diverse set of chromophores. The descriptors involved in the model relate rationally to the physical origin, because T_d is found to be determined by the following structural factors of chromophore molecules: size (which is reflected in our equation by VRD1), shape (accounted by VRD1 and X5sol), resonances (by EEig06r), and transfers of intramolecular charge (by JGI7 and JGI8). The R^2 of the correlation for Y_d values was 0.9642 and the MRE for the prediction of T_d was 4.46%. 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 the development of new NLO chromophores.

Acknowledgements This work was supported by the National Natural Science Foundation of China (50025309 and 90201016). Special thanks were given to Prof. Todeschini and other members in Milano Chemometerics and the QSAR Group for providing the Dragon package for use in this research.

References

1. Prasad PN, Williams DJ (1991) Introduction to nonlinear optical effects in molecules and polymers. Wiley, New York
2. Shi S (1994) Comtemp Phys 35:21–36
3. Zyss J (1994) Molecular nonlinear optics: materials, physics and devices. Academic, Boston
4. Burland DM, Miller RD, Reiser O, Twieg RJ, Walsh CA (1992) J Appl Phys 71:410–417
5. Burland DM, Miller RD, Walsh CA (1994) Chem Rev 94:31–75
6. Dalton LR, Harper AW, Ghosn R, Steier WH, Ziari M, Fetterman H, Shi Y, Mustacich RV, Jen AK-Y, Shea KJ (1995) Chem Mater 7:1060–1081
7. Pan Q, Fang C, Li F, Zhang Z, Qin Z, Wu X, Gu Q, Yu J (2002) Mater Res Bull 37:523–531
8. Marder SR, Beratan DN, Cheng LT (1991) Science 252:103–106
9. Moylan CR, Twieg RJ, Lee VY, Swanson SA, Betterton KM, Miller RD (1993) J Am Chem Soc 115:12599–12600
10. Moylan CR, Ermer S, Lovejoy SM, McComb I-H, Leung DS, Wortmann Rd, Krdmer P, Twieg RJ (1996) J Am Chem Soc 118:12950–12955
11. Kim O-K, Fort A, Barzoukas M, Blanchard-Descec M, Lehna J-M (1999) J Mater Chem 9:2227–2232
12. Cho BR, Park SB, Seung Jae Lee, Son KH, Lee SH, Lee M-J, Yoo J, Lee YK, Lee GJ, Kang TI, Cho M, Jeon S-J (2001) J Am Chem Soc 123:6421–6422
13. Oberg K, Berglund A, Edlund U, Eliasson B (2001) J Chem Inf Comput Sci 41:811–814
14. Zeng XD, Xu X, Wang BF, Wang BC (2004) Chin Chem Lett 15:753–756
15. Bicerano J (1996) Prediction of polymer properties. Marcel Dekker, New York
16. http://www.disat.unimib.it/chm/Dragon.htm
17. Jurs PC (1996) Computer software applications in chemistry. Wiley, New York
18. Rao VP, Jen AKY, Wong KY, Drost KJ (1993) J Chem Soc Chem Commun 14:1118–1120
19. Gilmour S, Marder SR, Perry JW, Cheng LT (1994) Adv Mater 6:494–496
20. Shu C-F, Wang Y-K (1998) J Mater Chem 8:833–835
21. Matsui M, Marui Y, Kushida M, Funabiki K, Muramatsu H, Shibata K, Hirota K, Hosoda M, Tai K (1998) Dyes Pigm 38:57–64
22. Mao SSH, Ra Y, Guo L, Zhang C, Dalton LR, Chen A, Garner S, Steier WH (1998) Chem Mater 10:146–155
23. Wang P, Zhu P, Wu W, Kang H, Ye C (1999) Phys Chem Chem Phys 1:3519–3525
24. Davey MH, Lee VY, Wu L-M, Moylan CR, Volksen W, Knoesen A, Miller RD, Marks TJ (2000) Chem Mater 12:1679–1693
25. Zhao L, Wang S, Chang J (2000) Chin J Org Chem 20:750–753
26. Ledouxa I, Zyssa J, Barni E, Barolo C, Diulgheroff N, Quagliotto P, Viscardi G (2000) Synth Met 115:213–217
27. Katti KV, Raghuraman K, Pillarsetty N, Karra SR, Gulotty RJ, Chartier MA, Langhoff CA (2002) Chem Mater 14:2436–2438
28. Pan Q, Fang C, Zhang Z, Qin Z, Li F, Gu Q, Xu X, Yu J (2003) Opt Mater 22:45–49
29. Spraul BK, Suresh S, Sassa T, Herranz MA, Echegoyen L, Wada T, Perahiaa D, Smith DW Jr (2004) Tetrahedron Lett 45:3253–3256
30. http://www.cambridgesoft.com/
31. Todeschini R, Consonni V (2000) Handbook of molecular descriptors. Wiley-VCH, Weinheim
32. http://www.spss.com/
33. Jobson JD (1991) Applied multivariate data analysis, regression and experimental design, vol 1. Springer, Berlin Heidelberg New York
34. Balaban AT, Ciubotariu D, Medeleanu M (1991) J Chem Inf Comput Sci 31:517–523

35. Kier LB, Hall LH (1986) Molecular connectivity in structure-activity analysis. RSP-Wiley, Chichester
36. Estrada E (1995) J Chem Inf Comput Sci 35:701–707
37. Estrada E (1995) J Chem Inf Comput Sci 35:31–33
38. Estrada E (1996) J Chem Inf Comput Sci 36:844–849
39. Estrada E, Ramirez A (1996) J Chem Inf Comput Sci 36:837–843
40. Estrada E (1997) J Chem Inf Comput Sci 37:320–328
41. Gilvez J, Garcia R, Salabert MT, Soler R (1994) J Chem Inf Comput Sci 34:520–525
42. Kim YS, Kim JH, Kim JS, No KT (2002) J Chem Inf Comput Sci 42:75–81
43. Yin S, Shuai Z, Wang Y (2003) J Chem Inf Comput Sci 43:970–977