J Mol Model (2005) 12: 24–33
DOI 10.1007/s00894-005-0282-5

ORIGINAL PAPER

Jie Xu · Biao Chen

Prediction of glass transition temperatures of OLED materials using topological indices

Received: 19 January 2005 / Accepted: 13 May 2005 / Published online: 16 August 2005

Abstract The QSPR study was performed between topological indices and glass transition temperatures (T_gs) of organic light-emitting diode materials based on a diverse set of 80 compounds. A five-parameter correlation with the squared correlation coefficient R^2 = 0.9304 and an average absolute error of 7.7 K was obtained through step-wise multi-linear regression analysis with leave-one-out cross-validation. The new model proposed is predictive and requires only topological indices in the calculations and has the advantage of the relative ease in calculating the descriptors, which makes it easier to apply. The predicted results of the new model are comparable to those of the existing equation by using the Comprehensives Descriptors for Structural and Statistical Analysis approach.

Keywords Glass transition temperature · OLED · QSPR · Topological index

Abbreviations LOO: Leave-one-out · MLRA: Multi-linear regression analysis · OLED: Organic light-emitting diodes · QSPR: Quantitative structure-property relationship

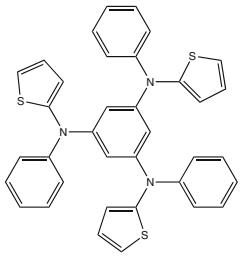
Introduction

Since the groundwork of Tang and Van Slyke [1], organic light-emitting diodes (OLED) have continued to be the subject of great interest because of their potential application to full-color flat-panel displays. Compared with liquid crystal displays, OLED displays possess wider view angles and a faster response time. In contrast to polymers, they are pure materials with well-defined molecular structures and definite molecular weights without any distribution. To obtain device quality, OLED materials must exhibit good thermal stability. The design and synthesis of thermally stable OLED materials is a key goal in this area of research [2–6]. With respect to the thermal stability, the glass transition temperature (T_g) is the most important factor for OLED materials. Therefore, a method for predicting the T_gs of OLED materials from their molecular structure would undoubtedly be valuable in the search for compounds and materials suitable for long-time use.

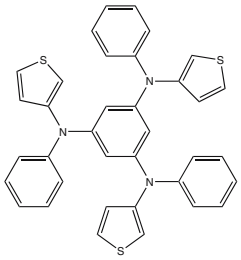
Alternatively, quantitative structure–property relationships (QSPR) provide a promising method for the estimation of the T_gs of OLED materials based on descriptors derived solely from the molecular structure to fit experimental data. The QSPR approach is based on the assumption that the variation of 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 only requires knowledge of the chemical structure and does not depend on any experimental properties. QSPR has been applied to the correlation of many diverse chemical, physical, biochemical, and pharmacological properties of chemical compounds successfully. In particular, Kim et al. [7] obtained a seven-parameter QSPR model with correlation coefficient R^2 = 0.989 for T_gs of small molecules by Genetic Algorithm (GA) and multi-linear regression. However, only 16 electroluminescent molecules, out of 81 molecules in total, are kept in the training set in their model. Recently, Yin et al. [8] used the COmprehensive DEscriptors for Structural and Statistical Analysis (CODESSA) program to develop a QSPR model (R^2 = 0.9270) for a set of T_g values of 73 OLED molecules, with six descriptors involved. These descriptors included constitutional, electrostatic, thermodynamic and quantum-chemical ones, which were derived from relatively sophisticated calculations.

J. Xu (✉) · B. Chen
Department of Polymer Science and Engineering,
University of Science and Technology of China,
Hefei, 230026, People's Republic of China
E-mail: xujie0@ustc.edu
Tel.: +86-551-3607874

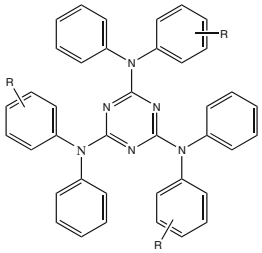
Fig. 1



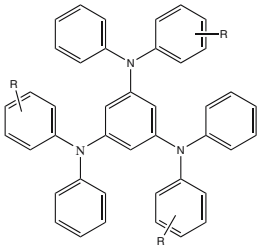
TPTAB1



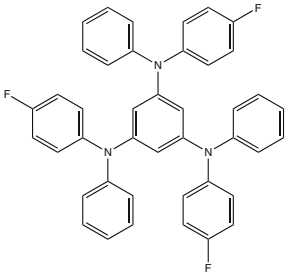
TPTAB2



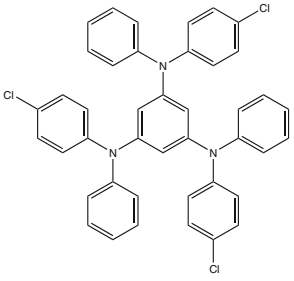
o-MTDATz:R=o-Me; m-MTDATz:R=m-Me



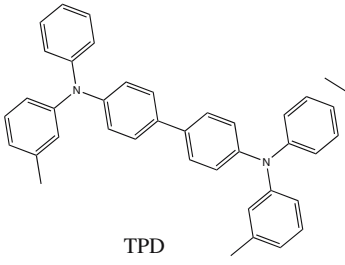
o-MTDAB:R=o-Me; p-MTDAB:R=p-Me



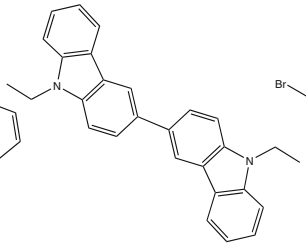
p-FTDAB



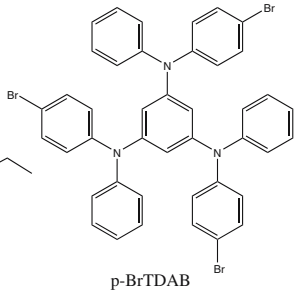
p-ClTDAB



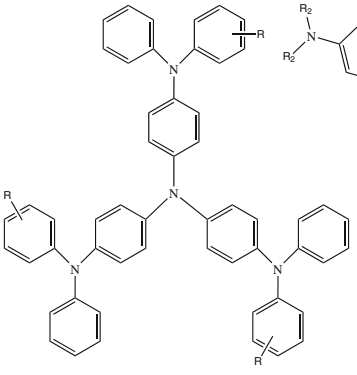
TPD



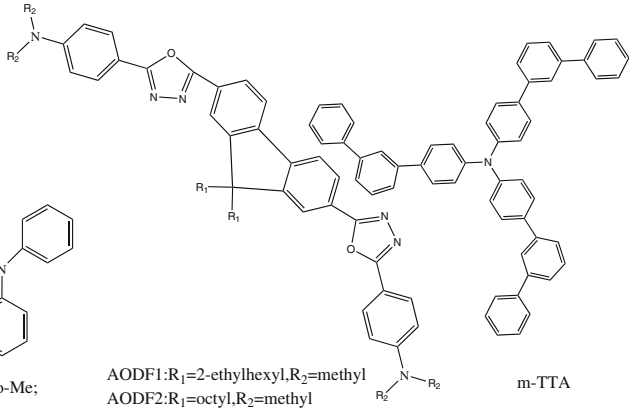
(EtCz)2



p-BrTDAB

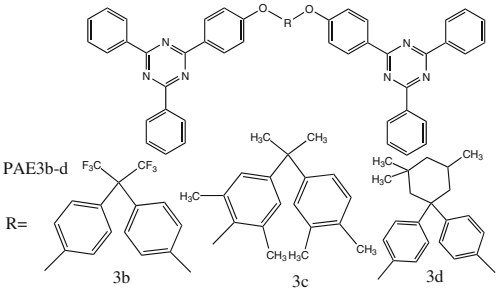


TDATA: R=H; o-MTDATA: R=o-Me;
m-MTDATA: R=m-Me



AODF1:R_1=2-ethylhexyl,R_2=methyl
AODF2:R_1=octyl,R_2=methyl

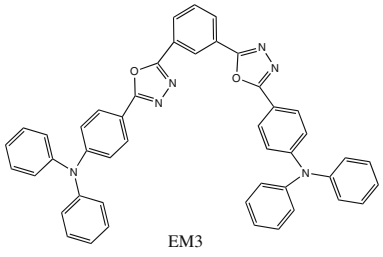
m-TTA



PAE3b-d

R=

3b 3c 3d



EM3

Fig. 1 (Contd.)

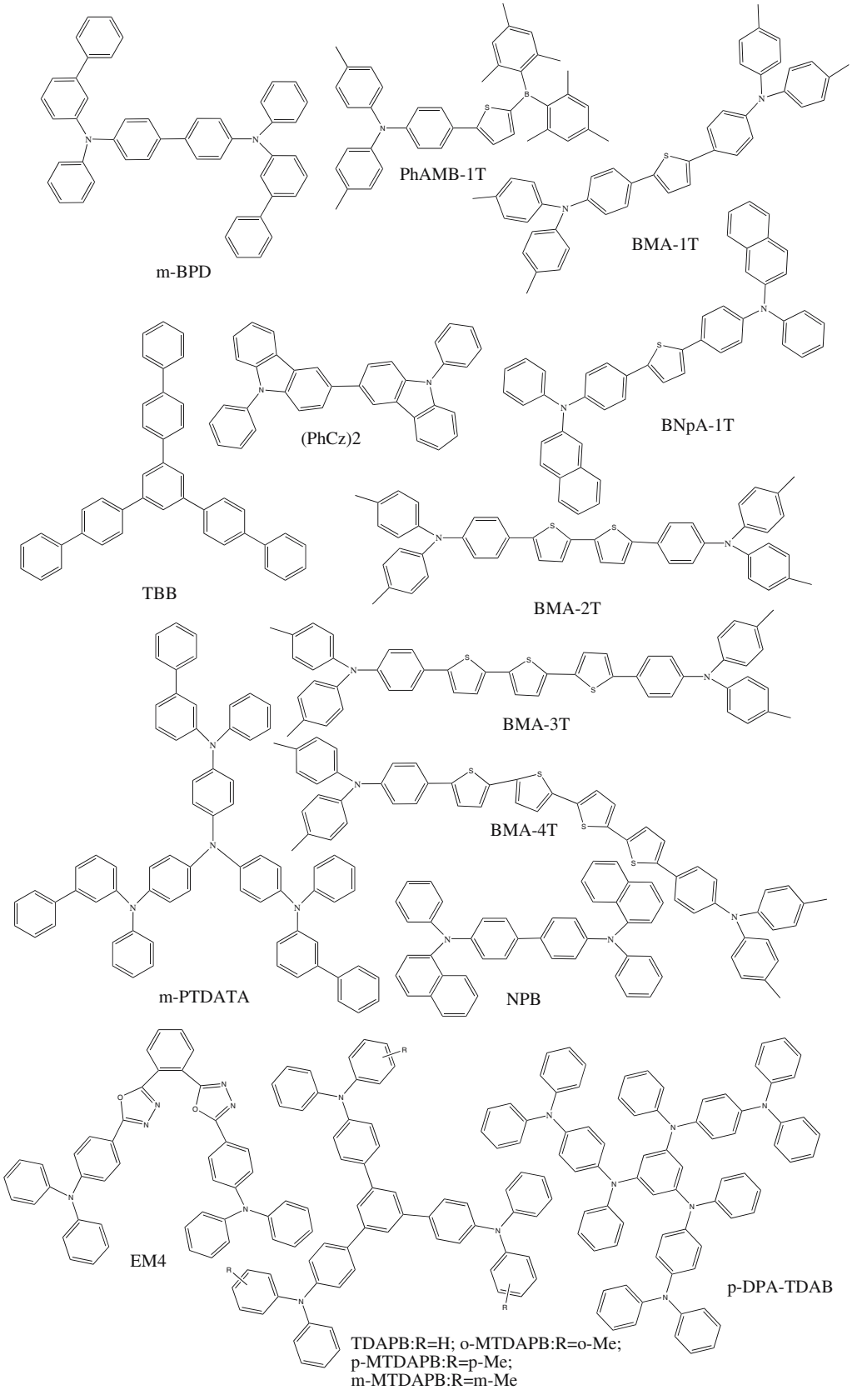
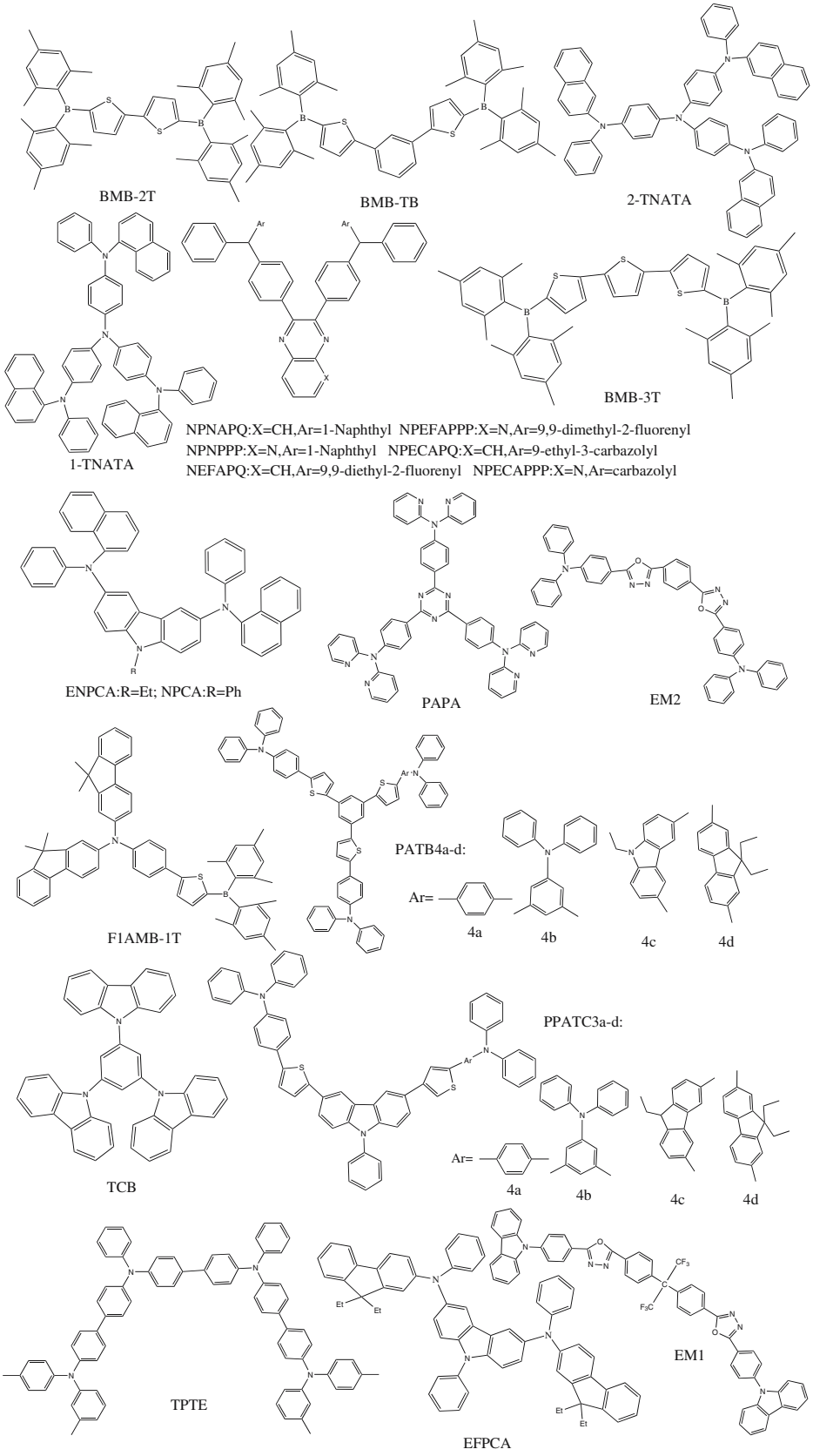

m-BPD
PhAMB-1T
BMA-1T
(PhCz)2
BNpA-1T
TBB
BMA-2T
BMA-3T
BMA-4T
m-PTDATA
NPB
EM4
p-DPA-TDAB
TDAPB:R=H; o-MTDAPB:R=o-Me;
p-MTDAPB:R=p-Me;
m-MTDAPB:R=m-Me

Fig. 1 (Contd.)


BMB-2T
BMB-TB
2-TNATA
BMB-3T
1-TNATA
NPNAPQ:X=CH,Ar=1-Naphthyl NPEFAPPP:X=N,Ar=9,9-dimethyl-2-fluorenyl
NPNPPP:X=N,Ar=1-Naphthyl NPECAPQ:X=CH,Ar=9-ethyl-3-carbazolyl
NEFAPQ:X=CH,Ar=9,9-diethyl-2-fluorenyl NPECAPPP:X=N,Ar=carbazolyl
ENPCA:R=Et; NPCA:R=Ph
PAPA
EM2
F1AMB-1T
PATB4a-d:
Ar=
4a
4b
4c
4d
TCB
PPATC3a-d:
Ar=
4a
4b
4c
4d
TPTE
EM1
EFPCA

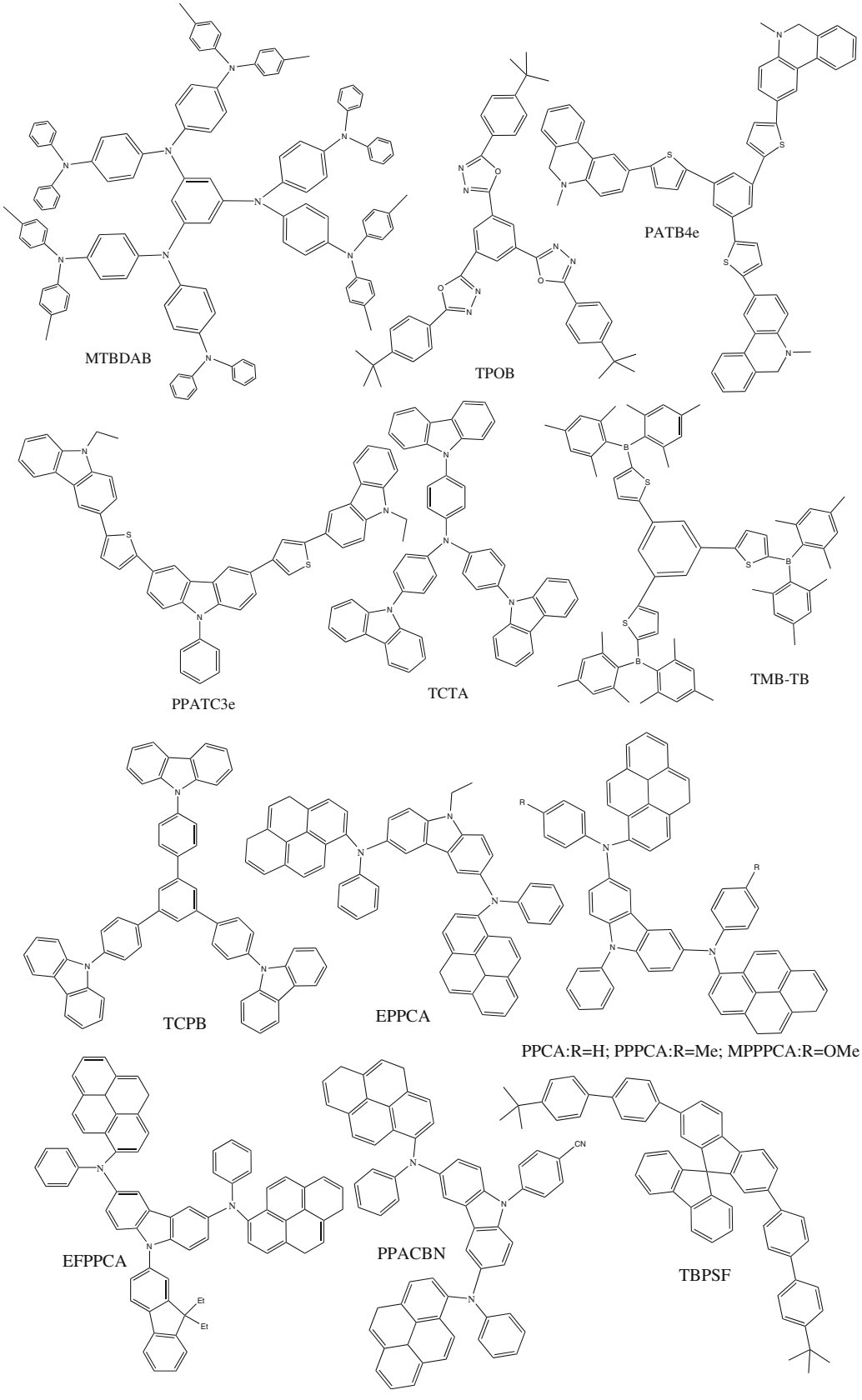

MTBDAB
TPOB
PATB4e
PPATC3e
TCTA
TMB-TB
TCPB
EPPCA
PPCA:R=H; PPPCA:R=Me; MPPPCA:R=OMe
EFPPCA
PPACBN
TBPSF

Fig. 1 (Contd.)

Topological descriptors, the commonly used molecular descriptors, have been used widely in the correlation of physicochemical properties of organic compounds. In chemical graph theory, molecular structures are normally represented as hydrogen-depleted graphs, whose vertices and edges act as atoms and covalent bonds, respectively. Chemical structural formulae can then be assimilated to undirected and finite multi-graphs with labeled vertices, commonly known as molecular graphs. Topological indices, also known as graph theoretical indices, are descriptors that characterize molecular graphs and contain a large amount of information about the molecule, including the numbers of hydrogen and non-hydrogen atoms bonded to each non-hydrogen atom, the details of the electronic structure of each atom, and the molecular structural features.

The QSPR models produced with only topological indices have advantages over models produced with other descriptors because of the relative ease of calculating descriptors, thus lowering computational cost and time [9, 10]. Further, because topological indices can be calculated solely from the molecular structure, models based on them are predictive. There have been numerous models developed with only topological indices [10–15]. Some of the recent studies are outlined here. García-Domenech and Julián-Ortiz [10] correlated the refractive indices and the glass transition temperatures of polymers with their topological indices; for the prediction of refractive indices, a ten-parameter correlation with R^2 of 0.962 was obtained. The T_gs were predicted through a model with R^2 of 0.894, consisting of ten indices. Zhong and Hu [15] have developed a QSPR model with R^2 of 0.885 to predict the aqueous solubility of organic compounds, with three topological indices involved.

The goal of the present study was to obtain an exclusively topological-index-based QSPR model that could predict T_g values for a diverse set of OLED materials. Molecular structures of the OLED materials were pre-optimized and only topological indices were calculated using TSAR 3.3. A five-parameter correlation

Table 1 Compounds used in the study with experimental and calculated glass transition temperature (T_g) (K) values

	Exp. T_g	Calc. T_g	ΔT_g[a]	Structure	Exp. T_g	Calc. T_g	ΔT_g[a]
TPTAB1	311	311.76	−0.76	NEFAPQ	389	394.02	−5.02
o-MTDAB	315	324.60	−9.60	EM3	391	387.72	3.28
m-MTDATz	315	341.36	−26.36	ENPCA	393	381.84	11.16
TPTAB2	319	314.74	4.26	PAPA	394	384.40	9.60
p-FTDAB	327	337.92	−10.92	TDAPB	394	372.99	21.01
o-MTDATz	328	325.12	2.88	EM2	395	387.39	7.61
p-MTDAB	328	335.21	−7.21	NPEFAPPP	395	395.26	−0.26
p-ClTDAB	337	335.511	1.489	NPCA	396	390.99	5.01
TPD	338	341.79	−3.79	F1AMB-1T	397	413.30	−16.30
(EtCz)2	343	350.16	−7.16	PATB4a	398	385.44	12.56
p-BrTDAB	345	331.56	13.44	TCB	399	404.98	−5.98
m-MTDATA	348	361.73	−13.73	NPNPPP	400	386.83	13.17
o-MTDATA	349	335.69	13.31	PPATC3a	401	400.31	0.69
AODF1	353	359.79	−6.79	TPTE	403	399.84	3.16
AODF2	353	356.10	−3.10	EFPCA	405	404.81	0.19
m-TTA	353	363.66	−10.66	PAE3d	406	424.49	−18.49
m-BPD	354	359.52	−5.52	EM1	407	425.44	−18.44
PhAMB-1T	357	351.11	5.89	MTBDAB	407	404.86	2.14
BMA-1T	359	355.43	3.57	TPOB	410	411.15	−1.15
TBB	361	352.19	8.81	PAE3c	412	397.32	14.68
TDATA	362	337.69	24.31	PATB4b	415	412.36	2.64
(PhCz)2	363	384.81	−21.81	PPATC3e	415	424.48	−9.48
BMA-2T	363	367.58	−4.58	NPECAPQ	415	404.34	10.66
BNpA-1T	364	374.36	−10.36	PATB4e	416	394.14	21.86
m-PTDATA	364	375.81	−11.81	PPATC3b	419	425.82	−6.82
BMA-3T	366	377.27	−11.27	PATB4d	423	433.25	−10.25
NPB	368	367.77	0.23	TCTA	423	419.30	3.70
BMA-4T	371	384.98	−13.98	NPECAPPP	425	424.91	0.09
EM4	372	381.44	−9.44	PPATC3d	429	434.16	−5.16
m-MTDAPB	378	383.42	−5.42	TMB-TB	433	430.61	2.39
p-DPA-TDAB	380	378.94	1.06	TCPB	445	436.66	8.34
BMB-2T	380	363.67	16.33	EPPCA	447	446.22	0.78
BMB-TB	382	386.69	−4.69	PATB4c	449	445.41	3.59
o-MTDAPB	382	371.60	10.40	PPATC3c	449	445.71	3.29
2-TNATA	383	385.26	−2.26	PPCA	453	453.27	−0.27
p-MTDAPB	383	378.43	4.57	MPPPCA	456	455.77	0.23
1-TNATA	386	388.75	−2.75	PPPCA	457	456.64	0.36
NPNAPQ	387	385.59	1.41	EFPPCA	458	465.31	−7.31
PAE3b	388	388.67	−0.67	PPACBN	467	455.81	11.19
BMB-3T	388	386.28	1.72	TBPSF	468	445.13	22.87

[a] $\Delta T_g = T_g$ (exp)− T_g (calc)

was then obtained through stepwise multi-linear regression analysis (MLRA) with leave-one-out (LOO) cross-validation. It can be used to predict the T_g values of the OLED materials, with the advantage of relative ease in calculating descriptors.

Materials and methods

Data set

The molecular structures of the OLED materials (Fig. 1) and the corresponding experimental data (Table 1) were taken from the article by Yin et al. [8]. A total of 80 compounds with extensive structural diversity were selected as the data set. The reported T_g values ranged from 311 to 468 K.

Descriptors generation and objective feature selection

The structures of all molecules were sketched on a PC using the ChemDraw program in the ChemOffice 7.0 software package [16]. Their three-dimensional (3D) structures were then generated, and their geometries were pre-optimized with the semi-empirical AM1 method using the Chem3D program in ChemOffice 7.0 to ensure that low energy conformations were obtained for each structure. The calculations were terminated if the energy gradient was smaller than 0.01 kcal mol^{-1} $Å^{-1}$. The output files were used for producing a set of topological descriptors using the TSAR 3.3 software [17], which calculates the following topological descriptors (Table 2): Wiener index [18], Balaban index [19–21], Randic–Kier–Hall subgraph connectivity indices [22–24], Kier–Hall electrotopological state indices [25, 26], Kier–Hall Kappa indices [27–29], and shape index [27–29]. The definitions of these topological indices are given in the Appendix.

A good correlation for structure and property should possess a high correlation coefficient R, low standard error s, and few variables. To this end, objective feature selection was done to remove those descriptors that provide minimal or redundant information. Descriptors with an essentially constant value for all structures were discarded. Pair-wise correlations of the remaining descriptors were examined to remove descriptors that are highly correlated with other descriptors. The reduction of the descriptors was also done to ensure that the ratio of descriptors to data set observations does not exceed 0.6, thereby reducing the risk of chance correlations during model development [30].

Table 2 Definitions of the topological indices used in this work

Index symbol	Definition
W	Wiener path number
J	Balaban index
${}^{m}\chi_p$	Path connectivity index of order m = 0–6
${}^{m}\chi^{v}{}_{p}$	Path valence connectivity index of order m = 0–6
${}^{m}\chi_c$	Cluster connectivity index of order m = 3–4
${}^{m}\chi^{v}{}_{c}$	Cluster valence connectivity index of order m = 3–4
${}^{4}\chi_{pc}$	Path-cluster connectivity index of order 4
${}^{4}\chi^{v}{}_{pc}$	Path-cluster valence connectivity index of order 4
${}^{m}\chi_r$	Ring connectivity index of order m = 3–6
${}^{m}\chi^{v}{}_{r}$	Ring valence connectivity of order m = 3–6
$S^{T}(i)$	Kier–Hall electrotopological state indices
${}^{m}\kappa$	Kappa index of order m = 1–3
${}^{m}\kappa_{\alpha}$	Kappa alpha index of order m = 1–3
Φ	Shape flexibility index

QSPR model development

To develop QSPR models, MLRA [31] with LOO cross-validation was applied to the data set. F-to-enter and F-to-leave values were 3 and 2, respectively. At first, multiple models were built using large subsets of randomly selected descriptors. Then the number of descriptors included in the procedure was gradually reduced by the selection of those that had higher t-values or appeared with higher frequency in previous models. Models with the number of descriptors no higher than eight, F-ratio higher than 40, cross-validated R^2 greater than 0.90, and correlation coefficient R higher than 0.95 between predicted and observed T_gs were validated using compounds of the data set. The following statistical characteristics of the models were used: correlation coefficient R, cross-validated R (denoted by R_{CV}) and coefficients of determination R^2 and $R^2{}_{CV}$ between calculated and experimental T_gs. Models were considered to have high predictive ability and good stability, if $R^2 > 0.90$ and $(R^2 - R^2{}_{CV})/R^2 < 0.1$.

Results and discussion

Step-wise MLRA with LOO cross-validation were used to select the descriptors for the best model and the quality was determined by examining the correlation coefficient R, the significance test F and the standard error s. Table 3 shows the best variable subsets of descriptors with different statistical parameters during the process of stepwise MLRA. It is noteworthy to point out that the most important descriptor seems to be the sixth-order path connectivity index ${}^{6}\chi_p$ because this descriptor appeared in all best variable subsets.

The best correlation model obtained for the entire data set of 80 OLED materials contains five descriptors (No. 9 in Table 3) and more descriptors evidently do not improve the regression results. The best five-parameter correlation equation is the following:

$$Tg(\mathrm{K}) = -131.568J + 14.778\,{}^{3}\chi^{v}_{c} - 18.242\,{}^{3}\chi^{v}_{p} + 10.977\,{}^{6}\chi_{p} + 16.652\,{}^{6}\chi^{v}_{p} + 469.005 \quad (1)$$

where $R^2 = 0.9304$, $R^2{}_{CV} = 0.9157$, $F = 197.643$, $s = 10.489$, $n = 80$.

Table 3 Results of step-wise multi-linear regression analysis (*MLRA*) with Leave-one-out (*LOO*) cross-validation for best correlation model

No.	Descriptor	*R*	*F*	*s*
1	$^{6}\chi_{p}$	0.7734	116.818	24.493
2	$^{6}\chi_{p}$, Φ	0.8745	124.689	18.923
3	$^{6}\chi_{p}$, Φ, *J*	0.9049	113.426	16.755
4	$^{6}\chi_{p}$, Φ, *J*, $^{4}\chi^{v}{}_{c}$	0.9199	100.883	15.628
5	$^{6}\chi_{p}$, Φ, *J*, $^{4}\chi^{v}{}_{c}$, $^{3}\chi^{v}{}_{p}$	0.9415	112.440	13.554
6	$^{6}\chi_{p}$, Φ, *J*, $^{4}\chi^{v}{}_{c}$, $^{3}\chi^{v}{}_{p}$, $^{3}\chi^{v}{}_{c}$	0.9564	114.087	12.421
7	$^{6}\chi_{p}$, Φ, *J*, $^{4}\chi^{v}{}_{c}$, $^{3}\chi^{v}{}_{p}$, $^{3}\chi^{v}{}_{c}$, $^{6}\chi^{v}{}_{p}$	0.9661	138.510	12.348
8	$^{6}\chi_{p}$, Φ, *J*, $^{3}\chi^{v}{}_{p}$, $^{3}\chi^{v}{}_{c}$, $^{6}\chi^{v}{}_{p}$	0.9660	163.743	10.523
9	*J*, $^{3}\chi^{v}{}_{c}$, $^{3}\chi^{v}{}_{p}$, $^{6}\chi_{p}$, $^{6}\chi^{v}{}_{p}$	0.9656	197.643	10.489

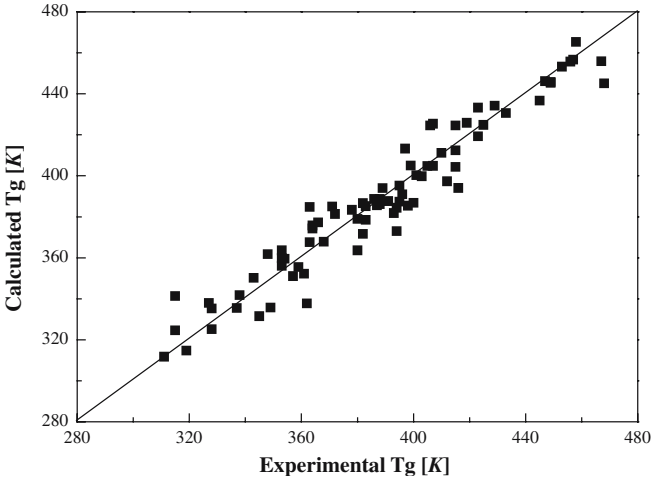


Fig. 2

The standard error is 10.5 K and the absolute average error (AEE) between the experimental and calculated T_g value is only 7.7 K for the 80 OLED materials in Table 1. The plot of calculated T_gs versus experimental T_gs for this model is shown in Fig. 2. The best five descriptors and the characteristics of these descriptors are shown in Table 4. The *t*-value indicates that the five parameters all are significant descriptors in the QSPR model. The cross-validated correlation coefficient (R^2_{CV} = 0.9157) shows the stability of this model. The data set of experimental T_gs was divided into two subsets according to their magnitude and the AAEs were essentially constant (7.9 and 7.6 K). This also indicates the stability of the model.

By interpreting the descriptors in the regression model, it is possible to gain some insights into factors that are likely to relate to the T_g values of OLED materials. According to Eq. 1, T_gs of OLED materials are influenced mainly by the molecular size and the molecular shape. The emergence of the sixth-order path and path–valence connectivity index $^{6}\chi_{p}$ and $^{6}\chi^{v}{}_{p}$ in the correlation equation is connected with the size of the molecule. Their positive signs indicate that the larger the OLED molecule is, the higher T_g is. The presence of the Balaban index *J* in the equation reflects the influence of the number of rings on the value of T_g. As shown in the definition of *J* in the Appendix, the index decreases with increasing number of rings. The negative sign of the Balaban index indicates that the more rings in the OLED molecule, the higher T_g is. The presence of the third-order cluster valence connectivity index $^{3}\chi^{v}{}_{c}$ shows the influence of the branching number on the value of T_g. The index increases with increased chain branching. The coefficient for the $^{3}\chi^{v}{}_{c}$ variable is positive, meaning that as branching increases, the T_g increases. The importance of the branch points in an aromatic ring to the T_g value is apparent due to the presence of the third-order path–valence connectivity index $^{3}\chi^{v}{}_{p}$. The $^{3}\chi^{v}{}_{p}$ index is larger for adjacent branch points than for separated branch points. Since the coefficient for the $^{3}\chi^{v}{}_{p}$ variable is negative, increased adjacency decreases the glass transition temperature.

The results obtained exclusively using topological descriptors herein are comparable in quality to those obtained by Yin et al. [8] for the glass transition temperatures of OLED materials by using the CODESSA approach (R^2 = 0.9304 vs. 0.9270, R^2_{CV} = 0.9157 vs. 0.9105, *F* = 197.64 vs. 139.61, *s* = 10.489 vs. 10.722, AAE = 7.7 K vs. 10.1 K). The CODESSA approach searches for linear correlations of a given property by using more than 600 descriptors of several classes: constitutional, topological, electrostatic, geometric, quantum-chemical, and thermodynamic ones. In Yin's model, the most significant descriptor is the relative number of rings (RNR), which encodes information about the chain stiffness of small molecules. It has very similar structural meaning as the Balaban index *J*, which

Table 4 Descriptors included in the best five-parameter model for the prediction of glass transition temperatures (T_gs)

Descriptor	Coefficient	SD	*t*-value	*t*-probability
Constant	469.005	13.367	35.086	0.000,000
J	−131.568	10.734	−12.257	0.000,000
$^{3}\chi^{v}{}_{c}$	14.778	2.005	7.370	0.000,000
$^{3}\chi^{v}{}_{p}$	−18.242	1.129	−16.161	0.000,000
$^{6}\chi_{p}$	10.977	0.739	14.852	0.000,000
$^{6}\chi^{v}{}_{p}$	16.652	1.948	8.549	0.000,000

is the second important descriptors in the present model. These facts confirm the potency of topological indices as a useful tool for the structural characterization and prediction of T_gs of OLED materials.

Conclusions

This work shows that the glass transition temperatures of OLED materials can be estimated with a simple correlation (R^2 = 0.9304 and AEE = 7.7 K) using only topological indices as input parameters. The proposed correlation contains five descriptors that indicate that T_gs of OLED materials are influenced mainly by the molecular size and the molecular shape. This model gives reasonable accuracy and it is predictive because topological indices can be calculated easily as long as the molecular structure of the compound concerned is known. Therefore, this model should be useful in development of new OLED materials.

Acknowledgments The authors gratefully wish to express their thanks to the reviewers for critically reviewing the manuscript and making important suggestions.

Appendix

The definitions of topological indices used in this work are shown below:

Wiener indices

The Wiener number, W, is the sum of distances in a molecular graph [18]. For a given connected molecular graph G

$$W = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} D_{ij}$$

where D_{ij} is a distance matrix of the shortest paths between any two vertices for N vertices. $D_{ij} = l_{ij}$ if $i \neq j$, otherwise equal to zero. l_{ij} is the shortest distance between vertices i and j.

Balaban indices

The Balaban index, J, is the average-distance sum connectivity [19–21]. For a given connected molecular graph G

$$J = \frac{M}{\mu + 1}\sum (D_i D_j)^{-0.5}$$

where M is the number of edges in G. μ denotes the ring number of G. In a polycyclic graph, μ is the minimum number of edges that must be removed before G becomes acyclic. $D_i = \sum_{j=1} D_{ij}$ and $D_{ij}$is defined as for the Wiener index.

Randic–Kier–Hall subgraph connectivity indices

The χ_t indices may be derived from the adjacency matrix [22–24] and they are defined as

$${}^{m}\chi_{\mathrm{t}} = \sum_{j=1}^{N_m} {}^{m}S_j$$

where m is the subgraph order, that is, the number of edges in the subgraph, N_m is the number of type t order m subgraphs within the whole graph, and ${}^{m}S_j$ is a factor defined for each subgraph as

$${}^{m}S_j = \prod_{i=1}^{m+1} (\delta_i)_j^{-1/2}$$

where j denotes the particular set of edges that constitutes the subgraph and δ_i is the degree of vertex i, that is, its number of edges.

Valence connectivity indices are defined similarly, substituting δ_i by δ_i^{v}, defined as

$$\delta_i^v = \frac{Z^v - h_i}{Z - Z^v - 1}$$

where Z is the atomic number of the atom i,Z^{v} the number of valence electrons, and h_i is the number of H atoms attached to it.

Kier–Hall electrotopological state indices

Kier and Hall [25, 26] developed electrotopological state indices (E-state indices), based on the electronegativity of an atom and its local topology. The E-state (S_i) of an atom i (non-hydrogens) is calculated by evaluating the intrinsic state value (I_i) and the perturbation arising from all the other skeletal atoms (ΔI_i):

$$S_i = (I_i + \Delta I_i)$$

where

$$I_i = \frac{[(2/N_i)^2\delta^{\mathrm{v}} + 1]}{\delta}, \qquad \Delta I_i = \sum_{j=1}^{N_i} (I_i - I_j)/r_{ij}^2$$

and: N_i is the principal quantum number of atom i,δ^{v} is the count of valence electrons in the skeleton; δ the count of s electrons in the skeleton (see Randic–Kier–Hall Subgraph Connectivity Indices, earlier), ΔI_i the perturbation on atom i by all other skeletal atoms j,r_{ij} is the number of atoms in the shortest path between atoms i and j (including i and j). Electrotopological state indices of the type of atoms are obtained by summing the electrotopological states for each present type of atoms in the molecule, S^{T} (i).

Kier–Hall Kappa indices

The Kappa indices are the basis of a method of molecular structure quantitation in which attributes of molecular shape are encoded into three indices (Kappa values) [27–29]. These Kappa values are derived from counts of one-bond, two-bond, and three-bond fragments, each count being made relative to fragment counts in reference structures which possess a maximum and minimum value for that number of atoms.

Shape flexibility index or Φ. Flexibility of a molecule is directly related to the degree of linearity and the presence of cycles and/or branching [27–29]. The Kappa alpha indices measure these factors while also taking the effects of atomic identities on shape into account. Hall and Kier found that by combining $^{1}\kappa_{\alpha}$ and $^{2}\kappa_{\alpha}$ indices, a further index Φ, which measured flexibility, could be defined:

$$\Phi = ({}^{1}\kappa_{\alpha} \bullet {}^{2}\kappa_{\alpha})/A$$

References

1. Tang C, Van Slyke S (1987) App Phys Lett 51:913–915
2. Wong K-T, Wang Z-J, Chien Y-Y, Wang C-L (2001) Org Lett 3:2285–2288
3. Yeh H-C, Lee R-H, Chan L-H, Lin T-YJ, Chen C-T, Balasubramaniam E, Tao Y-T (2001) Chem Mater 13:2788–2796
4. Wong K-T, Chien Y-Y, Chen R-T, Wang C-F, Lin Y-T, Chiang H-H, Hsieh P-Y, Wu C-C, Chou CH, Su YO, Lee G-H, Peng S-M (2002) J Am Chem Soc 124:11576–11577
5. Hughes G, Wang C, Batsanov AS, Fern M, Frank S (2003) Org Biomol Chem 1:3069–3077
6. Lee M-T, Yen C-K, Yang W-P, Chen H-H, Liao C-H, Tsai C-H, Chen CH (2004) Org Lett 6:1241–1244
7. Kim YS, Kim JH, Kim JS, No KT (2002) J Chem Inf Comput Sci 45:75–81
8. Yin S, Shuai Z, Wang Y (2003) J Chem Inf Comput Sci 43:970–977
9. Mattioni BE, Jurs PC (2002) J Chem Inf Comput Sci 42:232–240
10. García-Domenech R, Julián-Ortiz JVD (2002) J Phys Chem B 106:1501–1507
11. Kier LB, Hall LH (1976) Molecular connectivity in chemistry and drug research. Academic Press, New York
12. Kier LB, Hall LH (1986) Molecular connectivity in structure-activity analysis. Wiley, New York
13. Pogliani L (2000) Chem Rev 100:3827–3858
14. Zhong C, Yang C, Li Q (2002) Ind Eng Chem Res 41:2826–2833
15. Zhong C, Hu Q (2003) J Pharm Sci 92:2284–2294
16. ChemOffice 7.0 (2002) Inc Cambridgesoft, USA
17. TSAR 3.3 (2000) Oxford Molecular Ltd., USA
18. Wiener H (1947) J Am Chem Soc 69:17–20
19. Balaban AT (1982) Chem Phys Lett 89:399–404
20. Balaban AT (1983) Pure Appl Chem 55:199–206
21. Balaban AT (1986) Math Chem (MATCH) 21:115–122
22. Randic MJ (1975) J Am Chem Soc 97:6609–6615
23. Kier LB, Di Paolo T, Hall LH (1977) J Theor Biol 67:585–595
24. Kier LB, Hall LH (1983) J Pharm Sci 72:1170–1173
25. Kier LB, Hall LH (1990) Pharm Res 7:801–807
26. Kier LB, Hall LH (1999) Molecular structure description: the electrotopological state. Academic Press, San Diego
27. Kier LB (1985) Quant Struct-Act Relat 4:109–116
28. Hall LH, Kier LB (1992) In: Lipkowitz KB, Boyd DB (eds) Reviews in computational chemistry. Wiley, New York, pp 265–289
29. Kier LB, Hall LH (1999) In: Devillers J, Balaban AT (eds) Topological indices and related descriptors in QSAR and QSPR. Gordon and Breach Science Publishers, Amsterdam, pp 455–489
30. Topliss J, Edwards R (1979) J Med Chem 22:1238–1244
31. Jurs PC (1996) Computer software applications in chemistry. Wiley, New York