

The Use of Topological Indices in QSAR and QSPR Modeling

John C. Dearden

Abstract Topological indices (TIs) are numerical representations of the topology of a molecule, and are calculated from the heavy atom graphical depiction of the molecule. One of the first TIs was that of Wiener in 1947, who showed that his index correlated well with the boiling points of alkanes. There are now many different TIs available, and many of them are discussed in this chapter, with respect largely to their use as descriptors in QSAR/QSPR modeling. Three types in particular stand out, molecular connectivities developed by Randić and Kier and Hall, electrotopological state (E-state) values developed by Kier and Hall, and information content indices developed by Basak and co-workers. New TIs are still appearing, despite some criticism that there are already too many types of TI, that they are difficult of interpretation, and that they are inferior to physicochemical descriptors in modeling.

Keywords Wiener • Randić • Information content • Molecular connectivity • Electrotopological state • Biodescriptors • Inverse QSAR • Software • Hostility to topological indices

1 Introduction

1.1 What Is QSAR?

QSARs (quantitative structure-activity relationships) and QSPRs (quantitative structure-property relationships) are mathematical correlations between a specified biological activity or molecular property and one or more physicochemical and/or

J.C. Dearden (✉)

School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University,
Byrom Street, Liverpool L3 3AF, UK
e-mail: j.c.dearden@ljmu.ac.uk

© Springer International Publishing AG 2017

K. Roy (ed.), *Advances in QSAR Modeling*, Challenges and Advances

in Computational Chemistry and Physics 24, DOI 10.1007/978-3-319-56850-8_2

molecular structural properties, known as descriptors since they “describe” the activity or property under examination. A simple example is given by Eq. 1 (Cronin et al. 2002) for the inhibition of growth of the aquatic ciliate *Tetrahymena pyriformis* by phenols:

$$\begin{aligned}\log(1/IGC_{50}) &= 0.53 \log D - 0.96 LUMO - 0.58 \\ n = 160 \quad r^2 &= 0.81 \quad q^2 = 0.80 \quad s = 0.34 \quad F = 340\end{aligned}\tag{1}$$

where IGC_{50} = concentration of a substituted phenol required to inhibit growth by 50%, D = its distribution coefficient between 1-octanol and water buffered to pH 7.35, $LUMO$ = energy of the lowest unoccupied molecular orbital of the chemical (a measure of electrophilicity), n = number of chemicals used to develop the QSAR (termed the training set), r = correlation coefficient, q = cross-validated correlation coefficient, s = standard error of prediction by the QSAR, and F = Fisher statistic (variance ratio).

For those not familiar with QSAR, a brief explanation of Eq. 1 and its accompanying statistics is apposite. Activities and toxicities are almost invariably used in QSAR as the logarithm of the reciprocal concentration (or dose) to produce a required effect, for two reasons: (a) activities and toxicities can range over many orders of magnitude, so taking logarithms makes the numbers easier to handle, and (b) QSARs can be considered as modifications of the van't Hoff isotherm, which relates the free energy change in a process to the logarithm of the equilibrium or rate constant controlling the process; QSPRs in particular are sometimes referred to as linear free energy relationships (LFERs) (Wells 1968).

The statistical information provided gives an indication of the goodness-of-fit, robustness and predictive ability of the QSAR model. The coefficient of determination, r^2 , is a measure of how well the QSAR models the data; a value of 0.81 means that the model explains 81% of the variation in $\log IGC_{50}$. Considering that IGC_{50} is a measure of in vivo toxicity, that is a very good value; one only rarely finds r^2 values much greater than 0.8 when modeling in vivo data, because of inherent error in the data. q^2 is an internal cross-validated coefficient of determination, an indicator of how predictive the QSAR is—that is, how well it predicts IGC_{50} values for chemicals that were not used to develop the QSAR. This is done by removing one chemical from the training set, re-developing the QSAR without that chemical, and then using the re-developed QSAR to predict the IGC_{50} value of the removed chemical. That chemical is then returned to the training set, another chemical is removed and the process repeated until all chemicals have been removed in turn. A combined predictive indicator q^2 is then calculated. It should be noted that q^2 is not now considered to be a good indicator of predictivity (Golbraikh and Tropsha 2002). It is preferable to use chemicals that have not been used at all to develop a QSAR model, in order to test its predictivity.

Some years ago a set of principles, the OECD Principles for the Validation of (Q)SARs, was devised as guidance for the development and use of QSARs (OECD 2004):

A valid QSAR/QSPR should have:

- (1) *a defined endpoint;*
- (2) *an unambiguous algorithm;*
- (3) *a defined domain of applicability;*
- (4) *appropriate measures of goodness-of-fit, robustness and predictivity;*
- (5) *a mechanistic interpretation, if possible.*

The OECD report went on to say:

It is recognised that it is not always possible, from a scientific viewpoint, to provide a mechanistic interpretation of a given (Q)SAR (Principle 5)...The absence of a mechanistic interpretation for a model does not mean that a model is not potentially useful in the regulatory context. The intent of Principle 5 is not to reject models that have no apparent mechanistic basis, but to ensure that some consideration is given to the possibility of a mechanistic association between the descriptors used in a model and the endpoint being predicted, and to ensure that this association is documented.

It is important to note that “it is not always possible...to provide a mechanistic interpretation of a given (Q)SAR”. There are far too many examples in the literature of descriptors being wrongly interpreted in an attempt to provide a mechanistic interpretation. Johnson (2008) commented that “QSAR has devolved into a perfectly practiced art of logical fallacy: *cum hoc ergo propter hoc* (with this, therefore because of this)”.

1.2 What Are Topological Indices?

There are now thousands of physicochemical and structural descriptors available for use in QSAR/QSPR modeling. The vast majority of these are calculated values, since experimental measurement is time-consuming and expensive, whereas calculation is rapid and less expensive with the wide range of software now available for that purpose (see Table 1). Among these calculated values are the descriptors known as topological indices, which are graph invariants that encode the topology of molecules depicted as graphs (Devillers 1999a), usually without hydrogen atoms (i.e., hydrogen-suppressed graphs). In such graphs, atoms are termed ‘vertices’ and bonds are termed ‘edges’. The graphs are two-dimensional (2D), as shown in Fig. 1, and show the non-hydrogen atoms and their connections with each other (their connectivity).

In order to calculate a TI, typically the values of adjacent vertices, or some function of them such as square root or reciprocal, are multiplied, and then summed across all edges. So, for the 2-methylpentane molecule shown in Fig. 1, a simple TI would be $1 \times 3 + 1 \times 3 + 3 \times 2 + 2 \times 2 + 2 \times 1$, which is 18.

Many different types of topological indices are now available, through the manipulation of adjacency matrices and distance matrices (across multiple edges), including the use of graphs with more than one edge between at least one pair of

Table 1 Some software for calculation of topological indices

Software	Indices calculated ^a	Website (all accessed on 11 July 2016)
ADAPT	χ (0–7) ^b , kappa (1–3), E-state, Wiener	http://research.chem.psu.edu/pcjgroup/adapt.html
ADMET Predictor	E-state	http://www.simulations-plus.com
ADMEWORKS Predictor	χ (0–7), kappa (1–3), E-state, Wiener	http://www.fqs.pl
Bluedesc	χ , WHIM, autocorrelation	http://www.ra.cs.uni-tuebingen.de/software/bluedesc/welcome_e.html
ChemDes	χ , kappa, E-state, information content, WHIM, autocorrelation	http://www.scbdd.com/chemdes/
Chemistry Development Kit	χ (0–1), kappa, WHIM	http://www.opentox.org/dev/documentation/components/cdk
ChemProp	E-state	http://www.ufz.de/index.php?en=34593
CODESSA	χ , kappa, flexibility, Wiener, Balaban J, information content	http://www.semichem.com
CORINA Symphony	Autocorrelation	http://www.mn.am.com/products/corinasymphony
Dragon	χ , E-state, Randić, Zagreb, information content, ETA, autocorrelation	http://www.taletе.mi.it/
JOELib	E-state, kappa, autocorrelation, Zagreb	http://www.ra.cs.uni-tuebingen.de/software/joelib/index.html
MathChem	χ , Zagreb, Randić, Balaban J, Wiener	https://pypi.python.org/pypi/mathchem
MDL QSAR	χ (0–10), kappa (1–3), flexibility, Shannon, Wiener, Platt	https://www.mdl.com/products/predictive/qsar/index.jsp
Molconn-Z	χ (0–10), kappa (1–3), flexibility, Shannon, Wiener, Platt	https://www.edusoft-lc.com/molconn/
Mold ²	χ , flexibility, Zagreb, Randić, Balaban J, Wiener, autocorrelation, information content	https://www.fda.gov.ScienceResearch/BioinformaticsTools/Mold2
Molecular Modeling Pro	χ (0–4), kappa (2), E-state, Wiener	https://www.chemistry-software.com
MOE (Molecular Operating Environment)	χ (0–1), kappa (0–3), flexibility, E-state, Wiener, Balaban J, Zagreb	http://www.chemcomp.com
MOLE db	χ , E-state, Randić, Zagreb, information, ETA, autocorrelation, WHIM	http://michem.disat.unimib.it/mole_db/
PaDEL-Descriptor	χ , kappa (1–3), E-state, Wiener, Zagreb, WHIM	http://padel.nus.edu.sg/software/padeldescriptor

(continued)

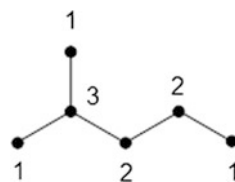
Table 1 (continued)

Software	Indices calculated ^a	Website (all accessed on 11 July 2016)
POLLY	information content (0–6), χ (0–6), Wiener	No website; copyright of University of Minnesota, 1988
PreADMET	χ , kappa, Wiener, Balaban J	https://preadmet.bmdrc.kr/
QSARPro	χ , kappa, information content	http://www.vlifesciences.com/products/QSARPro/Product_QSARpro.php
QuaSAR	χ (1, 2), kappa (1–3), flexibility, Wiener, Zagreb, Balaban J	http://www.chemcomp.com/journal/descr.htm
RDKit	χ (0–4), kappa (1–3), Balaban J	https://rdkit.readthedocs.io/en/latest/
SciQSAR	χ , kappa, E-state	http://www.pharmaceuticalonline.com/doc/sciqsar-2d-0001
T.E.S.T.	χ (0–10), kappa (1–3), E-state, information content, autocorrelation, Zagreb, Balaban J, Wiener	http://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test
TOPIX	χ (0–8), kappa (1–3), Randić, Wiener, Zagreb	http://www.lohninger.com/topix.html
VCCLAB Parameter Client	χ , kappa (1–3), flexibility, E-state, information content, Wiener, Randić, Balaban J, Zagreb, centric, autocorrelation, WHIM	http://www.vcclab.org/lab/pclient

^aSome software will calculate other topological descriptors in addition to those listed

^b χ (0–7) means χ values for path lengths 0 to 7; χ means no path lengths specified

Fig. 1 Hydrogen-suppressed graph of 2-methylpentane, with labelled vertices



adjacent vertices (i.e. taking account of double and triple bonds), termed multi-graphs, and weighted graphs, whereby the contributions of various edges and/or vertices are modified according to the relative importance of each to the TI. Many have proved useful in QSAR modeling (Netzeva 2004). An early example is the modeling of the potency of non-specific local anaesthetics by Kier et al. (1975), using an approach developed by Randić (1975). They obtained an excellent correlation of minimum blocking concentrations with what is now termed simple first order molecular connectivity, χ :

$$\begin{aligned}\log MBC &= 3.55 - 0.762\chi \\ n = 36 \quad r^2 &= 0.966 \quad s = 0.390\end{aligned}\tag{2}$$

The calculation of χ values is described in Sect. 16.

This chapter considers the usefulness of the main topological indices employed in QSAR and QSPR modeling. An early review of the subject is that of Balaban (1985). Those readers interested in the philosophy and theory of topological indices should consult Ivanciuc and Balaban (1999), Basak (2013a) and Roy et al. (2015). Rouvray and King (2002) and Todeschini and Consonni (2009) have presented and discussed, *inter alia*, a very wide range of topological indices. It should be noted that TIs are real numbers that represent aspects of molecular structure (Yilmaz and Götürk 2009), and thus are qualified for use as descriptors in QSAR/QSPR modeling.

1.3 The Value of Topological Indices

Topological indices have been shown to correlate well with numerous biological and physicochemical properties, suggesting that they are information-rich, and they are also generally quickly and readily calculated. They are therefore useful descriptors in QSARs and QSPRs that are used for predictive purposes, such as prediction of the toxicity of a chemical or the potency of a drug for future release in the market.

However, the term “descriptor” can be taken to relate not only to statistical description of the dependent variable, but also to physicochemical and/or structural description, implying that the descriptor(s) can yield information about the process (es) that control the magnitude of the dependent variable.

In connection with topological indices, Kubinyi (1993) forcefully pointed out that “in contrast to general recommendations on the selection of biologically meaningful parameters (descriptors), the physicochemical meaning of the topological parameters is never clear”. He later (Kubinyi 1997) described them as having a “hidden secret”. Livingstone (2000) has pointed out that with such large numbers of topological indices available for QSAR/QSPR use, there is a danger of chance correlations occurring. Lopez de Compadre et al. (1983) pointed out that there are dangers in their application to non-homologous series.

It has to be acknowledged that the inability of topological descriptors to allow much if any physicochemical interpretation is a grave drawback. Nevertheless, as Devillers (1999a) has pointed out, “these problems do not (mean) that topological indices must not be used in QSAR and QSPR studies. Indeed, they only show that, like all the other molecular descriptors, they have to be employed only in contexts for which they are suitable”. So long as this is recognised and acted upon, the use of topological indices is valid and valuable. Indeed, Randić et al. (2016) have made a

powerful case for the use of topological indices as true indicators of molecular structure in QSAR and QSPR modeling.

Of course, some TIs have proved more useful than others. In the present author's view, three types have proved especially valuable, namely (i) information content indices, developed by Basak and co-workers (see Sect. 7); (ii) molecular connectivities, devised initially by Randić as a branching index (see Sect. 13), and developed by Kier and Hall (see Sect. 14); and (iii) electrotopological state (E-state) indices, developed by Kier and Hall (see Sect. 19). The main types of TI in use today are discussed below in semi-chronological order of their introduction.

2 The Wiener Index

One of the first topological indices (Ivanciuc 2000) to be used in QSPR modeling is the Wiener index W (Wiener 1947), which gives an additive measure of the connections in a hydrogen-suppressed molecular graph. It is defined for hydrocarbons in terms of two variables, (a) the polarity number p , which is the number of pairs of carbon atoms that are separated by three carbon-carbon bonds, and (b) the path number w , calculated as follows: multiply the total number of carbon atoms on one side of any bond by those on the other side, and sum these for all bonds.

With this approach Wiener was able to predict the boiling points of a series of branched and straight chain paraffins to within an average of 1° , using Eq. 3:

$$\Delta t = (98/n^2) \Delta w + 5.5 \Delta p \quad (3)$$

where Δt = difference in boiling point between a straight and a branched chain isomer, and w and p = structural variables. Wiener later (Wiener 1948) used the same type of equation to model other physicochemical properties of isomeric alkanes. For example, for surface tension he found an average error of prediction of $0.13 \text{ dyne.cm}^{-1}$. The Wiener index has also been correlated with critical constants (Stiel and Thodos 1962), density and viscosity (Rouvray and Crafford 1976), and van der Waals surface area (Gutman and Körtvélyesi 1995).

Ivanciuc (2000) used two Wiener descriptors, weighted to account for the presence of heteroatoms and multiple bonds, along with $\log D$, to model the toxicity of 47 nitrobenzenes to *T. pyriformis*, with $r^2 = 0.875$ and $s = 0.250$, which is comparable to the model developed by Dearden et al. (1995) using physicochemical descriptors ($\log D$, $LUMO$ and modulus of change of charge on the nitro oxygen atom upon substitution), with $r^2 = 0.867$ and $s = 0.255$. Dearden et al. (1995) were able to make mechanistic interpretations of their results, concluding that the nitrobenzenes were behaving as pro-electrophiles, whilst Ivanciuc (2000) was unable to do so, as the Wiener descriptors yield little or no mechanistic information.

3 The Platt and Gordon-Scantlebury Indices

Platt (1947) devised a simple scheme to predict the physicochemical properties of alkanes, by summing the number of adjacent bonds for each atom. The index is not widely used, although Bharate and Singh (2011) found that it contributed to several good QSAR models of the anti-leishmanial effect of phloroglucinol-terpene adducts.

The Gordon-Scantlebury index (1964) is defined as the number of distinct ways that a sequence of three bonds can be overlapped on to the carbon skeleton of a molecule. Sabljic (1990) has pointed out that the index is equal to half the value of the Platt index. Like the Platt index, the Gordon-Scantlebury index is little used in QSAR. One instance of its use is an investigation of the antimycobacterial activity of alkenols (Gupta et al. 2005), although it did not compare well with other topological indices in that work.

4 The Hosoya Index

The Hosoya index Z is the number of sets of non-adjacent bonds in a molecule (Hosoya 1971). In other words, the Hosoya index of a hydrogen-suppressed graph is the total number of matchings within the graph, where a matching is a subset of edges that do not share a vertex. Like other topological indices, it gives a measure of molecular branching. Solomon et al. (2009) used it, along with a number of other descriptors, to model the inhibition of cholinesterase activity, although their best models included the Wiener index rather than the Hosoya index.

5 The Zagreb Indices

The first Zagreb index is calculated simply as the sum of the squares of the number of non-hydrogen bonds formed by each heavy atom (Gutman and Trinajstić 1972). A number of modifications of this were later developed, and Nikolić et al. (2003) have discussed these in detail. Whilst there has been much discussion on the derivation of Zagreb indices, there are relatively few publications that demonstrate their value in QSAR/QSPR modeling (Singh et al. 2014). Two such are that of Bajaj et al. (2005), who used refined Zagreb indices to model the anti-inflammatory activity of *N*-arylanthranilic acids, and that of Dureja et al. (2008), who found that the Zagreb topochemical indices M_1 and M_2 were valuable in modeling the fraction bound and clearance of cephalosporins in humans.

6 The Balaban J Index

The Balaban index (Balaban 1982), also called the average distance sum connectivity index, is computed as follows: the numbers of edges from atom i to all other atoms in a molecule are summed, and this procedure is repeated for all other atoms. The sums for adjacent atoms are then multiplied together, and the reciprocal square roots taken and summed. This number is then multiplied by $(B/(C + 1))$, where B = number of bonds in the molecule and C = number of rings, to give the Balaban index J . Unlike other topological indices, it does not increase rapidly with molecular size (Maran et al. 2010).

Mekenyan et al. (1987) found that J did not correlate well with a range of physicochemical properties or with acute toxicity of ethers. However, Thakur et al. (2004) found that the inhibition of carbonic anhydrase by sulphonamides was modeled well by J :

$$\begin{aligned}\log K_c &= 31.41 - 9.619 J \\ n &= 29 \quad r^2 = 0.910 \quad s = 0.429 \quad F = 274.2\end{aligned}\tag{4}$$

where K_c = inhibition constant.

7 Information Content Indices

Shannon (1948) was probably the first to study the science of information theory, which relates to molecular complexity (Mowshowitz 1968). However, the main proponents of this approach are undoubtedly Subhash Basak of the University of Minnesota and his co-workers (Basak 1999, 2013b and references cited therein).

There are four main types of information indices: mean, total, complementary and structural information content (Basak 1999). The reader is referred to the publications of Basak (1999, 2013b) for details of the calculation of these indices. Like most topological indices, information content indices appear to work best with homologous series, as with the examples cited by Basak (1987), such as the anesthetic potency (AD_{50}) of barbiturates;

$$\begin{aligned}AD_{50} &= 0.33TIC_1 - 0.002(TIC_1)^2 - 18.50 \\ n &= 13 \quad r^2 = 0.98 \quad s = 0.06\end{aligned}\tag{5}$$

where TIC_1 = first-order total information content.

This is statistically a slightly better model than that obtained using $\log P$ (octanol-water partition coefficient):

$$\begin{aligned}AD_{50} &= 1.58 \log P - 0.44(\log P)^2 + 1.93 \\ n = 13 \quad r^2 &= 0.94 \quad s = 0.10\end{aligned}\tag{6}$$

A combination of information indices and other descriptors can yield good QSAR models for diverse data sets. For example, for acute toxicity to fathead minnow of 69 diverse benzene derivatives (Gute and Basak 1997), a combination of topostructural and topochemical indices yielded $r^2 = 0.783$ and $s = 0.36$. When geometric and quantum chemical descriptors were added, the correlation improved: $r^2 = 0.863$, $s = 0.30$. However, geometric and quantum chemical descriptors alone did not yield good models.

8 Autocorrelation Descriptors

The autocorrelation approach, first introduced by Moreau and Broto (1980), derives molecular descriptors encoding various physicochemical or structural properties from the molecular graphs of the organic chemicals being studied (Devillers 1999b). The procedure is as follows:

- (1) The shortest interatomic distances, expressed as number of bonds, between each pair of atoms i and j are calculated.
- (2) An appropriate physicochemical or structural property is chosen, and the autocorrelation vector is calculated as the sum of the products of the atomic contributions to that property for each distance between the different atoms.

Many physicochemical properties have been used in QSARs and QSPRs involving this approach (Devillers 1999b). González et al. (2006) used masses, electronegativities and van der Waals volumes to model the inhibitory activity of cytokinin-derived cyclin-dependent kinase inhibitors.

Abreu et al. (2009) used a combination of autocorrelation descriptors and radial distribution descriptors to model the radical scavenging activity of benzo[*b*]thiophenes. In a comparative assessment of 2D autocorrelation, CoMFA and CoMSIA modeling of protein tyrosine kinase inhibition, Caballero et al. (2008) found that CoMSIA performed best.

9 WHIM Descriptors

WHIM (Weighted Holistic Invariant Molecular) descriptors are geometrical descriptors based on statistical indices calculated on the projections of atoms along principal axes (Todeschini et al. 1994). They are able to capture relevant molecular

information regarding molecular size, shape, symmetry and atom distribution with respect to invariant reference frames. In the WHIM approach a molecule is considered as a configuration of points (the atoms) in the three-dimensional space defined by the Cartesian axes. Projections of the atoms along each principal axis are made, and their distributions around the geometric centre are evaluated.

Todeschini and Gramatica (1997) found that WHIM descriptors performed very well in the prediction of a number of physicochemical properties of chlorophenols and of their aquatic toxicity to a range of species. Vlaia et al. (2009) obtained excellent correlations of WHIM descriptors with the toxicity of 48 aliphatic esters to the aquatic ciliate *Tetrahymena pyriformis*. Tong et al. (2008) used the vectors of principal component scores of WHIM indices of peptide analogues to model properties such as bitter taste ($n = 48$, $r^2 = 0.873$, RMSE (root mean square error) = 0.225) and bactericidal activity ($n = 12$, $r^2 = 0.997$, RMSE = 0.133).

10 Topochemical Atom Indices

Topochemically Arrived Unique (TAU) indices, developed by Pal et al. (1988), take account of the chemistry of the atomic core and valence electronic environment of atoms. A detailed explanation of their derivation has been given by Roy and Saha (2003), who used them to model the aqueous solubility of 193 diverse acyclic compounds, with excellent results ($r^2 = 0.946$, $s = 0.735$). Roy and Ghosh (2003) then extended the scope of the TAU scheme by redefining its basic parameters and introducing a novel Extended Topochemical Atom (ETA) formalism. They showed that their new ETA indices could model the toxicities of 50 substituted phenols to *Tetrahymena pyriformis* very well ($r^2 = 0.948$, $q^2 = 0.936$, $s = 0.161$, $F = 159.1$). They also found (Roy and Ghosh 2004) a good correlation of ETA indices with acute toxicity of substituted benzenes to the guppy ($r^2 = 0.885$, $q^2 = 0.865$, $s = 0.230$, $F = 92.6$), and a good correlation (Roy and Ghosh 2009) of a combination of 15 ETA and non-ETA topological descriptors with the toxicity of 288 diverse aromatic compounds to *Tetrahymena pyriformis* ($r^2 = 0.854$, $q^2 = 0.821$, is not given, $F = 106.3$). It may be noted that the use of a large number of descriptors in a QSAR is not recommended (Aptula et al. 2005) as it makes interpretation difficult.

11 The Centric Index

The centric index C developed by Balaban (1979) reflects molecular shape. It uses a procedure known as pruning partition of terminal atoms, and is calculated as follows:

$$C = \Sigma(a_i)^2 \quad (7)$$

where a_i = number of atoms deleted in step i .

The index has not been widely used in QSAR and QSPR investigations. Two studies that employed it (Jalali-Heravi and Asadollahi-Baboli 2008; Noorizadeh et al. 2011) found it not to feature in their best models.

12 Triplet Indices

Filip et al. (1987) introduced a new approach for obtaining graph invariants with very high discrimination ability, called triplet indices. Local vertex invariants (LOVIs) are assembled into a triplet TI, based on one of several operations such as: (i) summation; (ii) summation of squares; (iii) summation of square roots, and so on (Basak et al. 2000). Filip et al. (1987) showed that triplet indices correlated well with physicochemical properties such as boiling points, and NMR chemical shifts.

13 The Randić Index

Randić (1975) devised a topological index to characterize branching in alkanes. In Fig. 2 are depicted the hydrogen-suppressed graphs of three isomeric hexanes, namely *n*-hexane, 2-methylpentane and 3-methylpentane. The branching index (BI) for each is calculated by multiplying the number of non-hydrogen bonds made by each atom with the number on an adjacent atom, taking the reciprocal square root of the product, then summing across all non-hydrogen atoms.

Hence, for *n*-hexane the BI is $1/\sqrt{2} + 1/\sqrt{4} + 1/\sqrt{4} + 1/\sqrt{4} + 1/\sqrt{2}$, or 2.914. The BIs for 2-methylpentane and 3-methylpentane are 2.770 and 2.808 respectively. Clearly the Randić index can readily differentiate between alkane isomers.

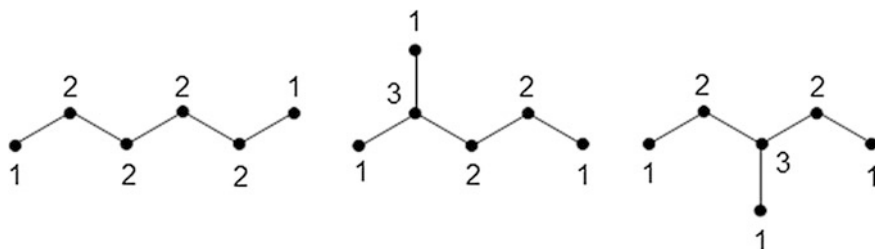


Fig. 2 Hydrogen-suppressed graphs of three isomeric hexane molecules

This was an important development, and was recognised as such by Lemont (Monty) Kier and Lowell Hall, who proceeded to develop Randić's concept and widen its applications (see Sect. 16). Randić himself later (Randić 2001) generously acknowledged the significant contributions of Kier and Hall, quoting Wilson (1952): "Every once in a while some new theory or a new experimental method or apparatus makes it possible to enter a new domain. Sometimes it is obvious to all that this opportunity has arisen, but in other cases recognition of the opportunity requires more imagination".

14 Molecular Connectivity Indices

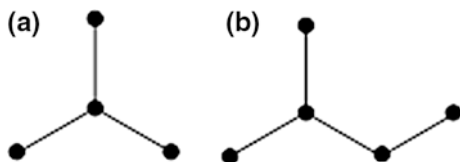
Kier and Hall, discerning the potential of the Randić index, collaborated with Randić (Kier et al. 1975) to show that his branching index could be applied to compounds other than alkanes, and could be used in QSPR and QSAR modeling. They demonstrated that the cavity surface area of 69 alcohols and hydrocarbons was well modelled by BI, which they chose to call the molecular connectivity index, χ ($r^2 = 0.956$, $s = 11.2$). Such a correlation is not unexpected, since χ clearly increases with molecular size. What is much more interesting is that they also correlated a biological activity (minimum blocking concentration of non-specific local anaesthetics) with χ for a chemically diverse set of 36 chemicals ($r^2 = 0.966$, $s = 0.390$).

It was also realised (Murray et al. 1975) that for compounds with π -bonds, better correlations could be achieved if a π -bond was regarded as two bonds, so that, for example, $\text{CH}_2 =$ has a vertex value (connectivity) of 2. Kier et al. (1976) also introduced the concept of an expanded series of the molecular connectivity index χ , involving calculation across more than one bond. For example, for isopentane, the first-order χ value, designated as $^1\chi$, is $1/\sqrt{3} + 1/\sqrt{3} + 1/\sqrt{6} + 1/\sqrt{2}$, or 2.270. The second-order χ value, calculated by multiplying across two bonds, and designated as $^2\chi$, is $1/\sqrt{1} \times 3 \times 1 + 1/\sqrt{1} \times 3 \times 2 + 1/\sqrt{1} \times 3 \times 2 + 1/\sqrt{3} \times 2 \times 1$, or 1.558. The third-order χ value, calculated by multiplying across three bonds, and designated as $^3\chi$, is $1/\sqrt{1} \times 3 \times 2 \times 1 + 1/\sqrt{1} \times 3 \times 2 \times 1$, or 0.816. Higher-order χ values are calculated similarly, and the zero-order χ value is calculated as the sum of the reciprocal square roots of the vertex values.

A further step (Kier and Hall 1976a) involved the specific treatment of heteroatoms. The vertex value δ of an atom is equivalent to the number of valence electrons minus the number of hydrogen atoms bonded to it; for example, for the N atom of NH_2 , $\delta = 3$, and for the N atom of NH , $\delta = 4$. However, this does not work for the halogens. To circumvent this problem, the δ values for halogen atoms were derived from modeling of molar refraction data, yielding δ values of: fluorine -20, chlorine 0.690, bromine 0.254, iodine 0.085.

When χ values corrected for unsaturation and heteroatoms are used, they are written as χ^v . For example, the simple (uncorrected) $^1\chi$ value for vinyl chloride,

Fig. 3 **a** 3rd order cluster;
b 4th order path-cluster



$\text{CH}_2 = \text{CHCl}$, is $1/\sqrt{2} + 1/\sqrt{2}$, or 1.414. The valence-corrected ${}^1\chi^v$ value is $1/\sqrt{2} \times 3 + 1/\sqrt{3} \times 0.690$, or 1.103.

Also in 1976 Kier and Hall (1976b) published their first book on molecular connectivity.

When branching occurs in a molecule, the atoms at and around the branch are termed a cluster (Fig. 3a) or a path-cluster (Fig. 3b). Clearly their molecular connectivity terms (${}^3\chi_c$ and ${}^4\chi_{pc}$) describe local structural properties. Kier et al. (1977) found them useful in modelling odorants (Eq. 8) and Sabljic and Protić-Sabljic (1983) used them to model properties of branched alcohols.

$$\begin{aligned}\text{Odor similarity} &= 7.47 - 1.84^2\chi + 1.34^3\chi_c \\ n = 15 \quad r^2 &= 0.848 \quad s = 0.395\end{aligned}\quad (8)$$

Ring (termed chain) molecular connectivities describe the types of rings and their substitution patterns in a molecule (Kier and Hall 1986). Sabljic (1985) found that a chain term was required to model chromatographic retention indices of chlorinated benzenes on a polar stationary phase:

$$\begin{aligned}I^{\text{CW20M}} &= 226.8^3\chi + 1588.0^7\chi_{\text{CH}} + 649.1 \\ n = 13 \quad r^2 &= 0.996 \quad s \text{ not given} \quad F = 1347\end{aligned}\quad (9)$$

A differential molecular connectivity index, $\Delta\chi$, was introduced in Kier and Hall (1991), defined as the difference between the simple and valence connectivity indices of the same order. The information encoded by this differential index is largely electronic. For example, Kier and Hall (1991) found that the ionization potentials (IP) of amines, alcohols and ethers were well modeled by two $\Delta\chi$ values:

$$\begin{aligned}\text{IP} &= 5.01 \Delta^0\chi + 5.17 \Delta^1\chi + 5.34 \\ n = 24 \quad r^2 &= 0.912 \quad s = 0.30 \quad F = 109\end{aligned}\quad (10)$$

Kier and Hall have utilized molecular connectivities to model a wide range of physicochemical and biological endpoints (Hall and Kier 1999a), from aqueous solubility (Hall et al. 1975) to muscarinic receptor affinity (Kier and Hall 1978) to fish toxicity (Hall et al. 1989).

Hall and Kier (1999a) have also pointed out that molecular connectivities, together with kappa and E-state indices (see Sects. 17 and 21), have been utilized in

database characterization (Cummins et al. 1996) and combinatorial library design (Zheng et al. 1998a, b).

It is not surprising that there is considerable collinearity of χ values, especially amongst the lower order values. It is important to eliminate collinearity of descriptors in a QSAR model, otherwise distortion of the statistics can occur (Dearden et al. 2009) and mechanistic interpretation is difficult. Murcia-Soler et al. (2001) used χ values to model the anti-hyperglycemic effect and other properties of sulfonyleurea drugs. They modeled the plasma protein binding of those drugs with three molecular connectivities, namely ${}^0\chi^v$, ${}^1\chi$ and ${}^1\chi^v$, all of which are highly correlated with each other ($r > 0.99$). Lu et al. (1999) modeled the fish bioconcentration factor (BCF) of organic pollutants, and reported the following model:

$$\log \text{BCF} = 0.770 + 0.757{}^0\chi^v - 2.650{}^1\chi + 3.372{}^2\chi - 1.186{}^2\chi^v - 1.807{}^3\chi_c \quad (11)$$

$$n = 80 \quad r^2 = 0.907 \quad s = 0.364$$

They did not give the χ values of the compounds, but by comparison with the Murcia-Soler data above it is clear that at least one pair of descriptors (${}^0\chi^v$ and ${}^1\chi$) in Eq. 11 must be very highly correlated.

Randić (2001) took a different view of collinearity. He stated: “Descriptors that show high collinearity with already selected descriptors are often eliminated from structure-property-activity studies. They should not be. The *only* useful criterion for discarding a descriptor is its inability to reduce the standard error of the regression. For example, in several applications of connectivity indices, the second order connectivity index ${}^2\chi$ has been discarded because...it shows close parallelism to the connectivity index ${}^1\chi$. But... ${}^2\chi$, despite its parallelism to ${}^1\chi$, also *complements* it. That is, a part of ${}^2\chi$ which is different from ${}^1\chi$ (and which may be small) suffices to produce satisfactory regression”. Randić (2001) pointed out that a referee disagreed with his view, stating that “such a model is generally not predictive, that is, when new compounds are predicted, their presence essentially alters the interrelation between the two descriptors, ${}^1\chi$ and ${}^2\chi$ in this example. Often when models using inter-correlated variables are used, they do not produce good validation statistics”. This is confirmed by, for example, Livingstone (1995) and Hansch et al. (1998). The latter authors pointed out that a QSAR developed by Ribo and Kaiser (1984) for the toxicity of chloroanilines to *Photobacterium phosphoreum* (now called *Vibrio fischeri*) contained two highly correlated terms, ClogP and the Hammett constant σ :

$$\log 1/C = 1.25(\pm 0.49) \text{ClogP} - 1.45(\pm 1.1)\sigma + 2.01(\pm 0.70) \quad (12)$$

$$n = 14 \quad r^2 = 0.917$$

The correlation between ClogP and σ is $r^2 = 0.946$, and the σ term has low significance, as can be seen from its standard error's being high relative to its coefficient.

The present author remains of the belief that it is probably better not to use highly correlated descriptors in a QSAR/QSPR. The subject is, however, worthy of further investigation. Hollas et al. (2005) have reported that by slight modification of TIs such as molecular connectivities, the Platt number and Zagreb indices, their mutual correlation can be reduced or completely eliminated.

Another point that perhaps requires further examination is just what is meant by “highly correlated”. There does not appear to be any definitive value, at least so far as QSAR is concerned, and a wide range of values are in use. The present author’s very subjective choice is $r^2 \geq 0.8$. Gramatica et al. (2007) used $r^2 \geq 0.98$, which aligns with the view of Randić (2001). On the other hand, r values as low as 0.2 have been used as a cut-off point (Randić 2015).

It has already been mentioned that molecular connectivities, like other topological descriptors, are difficult of interpretation. Nonetheless a number of attempts have been made to do so. Kier and Hall (1976b, 1986) pointed out that there are five general categories of molecular structure described by χ indices: (i) degree of branching, (ii) variable branching pattern, (iii) position and effect of heteroatoms, (iv) adjacency patterns, and (v) degrees of cyclicity. Kier and Hall (2000, 2001) also showed that χ indices represent the numerical possibilities of a molecule encountering another identical molecule. By converting bond accessibility into a cellular automata rule for 38 alkanes, and running the dynamics, they showed that the number of cell encounters correlated better ($r^2 = 0.991$) than did ${}^1\chi$ ($r^2 = 0.984$) with the boiling points of the alkanes.

Randić and co-workers (Randić and Zupan 2001; Randić et al. 2001) considered the interpretation of several topological indices. They pointed out that the paucity of papers on the subject suggested that the interpretation of topological indices may be rather difficult. Nonetheless they attempted to do so, and commented that the fact that peripheral bonds make larger contributions to ${}^1\chi$ than do inside bonds indicates their contribution to molecular surface area, which is a measure of molecular size.

Estrada (2002) also identified χ indices as components of molecular accessibility. He interpreted the δ values (inverse square roots of the vertex degree (number of non-hydrogen bonds formed)) as the length of the arc in the van der Waals circumference accessible from outside. Then, for a ${}^2\chi$ index (i.e. over 3 atoms), the 3 δ values are multiplied together to give a molecular accessibility volume.

In a principal component (PC) analysis of 108 *n*-alkanes and polychlorinated biphenyls, Burkhard et al. (1983) found that three PCs accounted for 98% of the variance in the data set, and that those PCs were associated with, respectively, (i) degree of branching, (ii) molecular size or bulkiness, and (iii) structural flexibility.

Dearden et al. (1988) looked at the correlations between a range of χ values of 59 substituents attached to a benzene ring and 54 non- χ properties of each of the substituents. In general, they found that path connectivity terms, of whatever order (≤ 6) and whether simple or valence-corrected, model predominantly bulk volume.

The $(\chi - \chi^v)$ terms did not appear to model any of the 54 non- χ properties. Other ad hoc comments have occasionally appeared in the literature; for example, Krishnasamy et al. (2008) stated that $^4\chi_p^v$ highlights the role of molecular surface. Stankevich et al. (1995) reported that χ indices for conjugated hydrocarbons correlated with the Hamiltonian function describing the π -electron properties of the compounds.

There have also been several attempts to “correct” χ indices. For example Li et al. (2003) developed a novel valence χ value and found it better ($r^2 = 0.939$) than the standard χ value ($r^2 = 0.889$) for the prediction of aqueous solubility of a diverse group of 36 organic compounds. Zhang et al. (2005) also used the same novel valence χ value, together with several quantum-chemical descriptors, in the modeling of corrosion inhibitory activity of 34 compounds such as imidazoles and imidazolines.

Dearden et al. (2004) devised an approach to improve the correlation of χ values with hydrophobicity by subtracting, instead of adding, the bond contributions (δ values) for bonds where one of the atoms is a heteroatom other than halogen, to give a $^1\chi^p$ value. For example, for a set of 23 diverse substituents, the correlation between $^1\chi$ and π (the hydrophobic substituent constant) was $r^2 = 0.123$, whilst that between $^1\chi^p$ and π was $r^2 = 0.771$.

15 Kappa Indices

Kier (1985) devised a numerical index (kappa, κ) of molecular shape from the hydrogen-suppressed graph of a molecule. It is based on the count of 2-bond fragments in a graph relative to the maximum number possible (if the molecule is star-shaped) and the minimum number in the isomeric linear graph. He showed that the sweet taste potency of 14 nitro- and cyano-anilines was modelled better ($r^2 = 0.852$, $s = 0.30$) with $^2\kappa$ and $(^2\kappa)^2$ than was found by Iwamura (1980) using a Verloop steric constant ($r^2 = 0.810$, $s = 0.32$). Kier later (Kier 1986a) introduced different orders of kappa indices and (Kier 1986b) a modification term α for non-carbon atoms.

Solomon et al. (2009) used a first-order κ index in a good 5-descriptor QSAR to model the butylcholinesterase inhibition of 59 *N*-aryl derivatives ($r^2 = 0.884$).

16 Flexibility Indices

Almost all organic molecules are flexible, and flexibility often plays an important part in chemical reactions, and in xenobiotic transport and receptor binding within an organism (Luisi 1977). Kappa indices were used by Kier (1989) to develop a

molecular flexibility index. The heteroatom-weighted kappa indices ${}^1\kappa_\alpha$ and ${}^2\kappa_\alpha$ are obtained, and the flexibility index Φ is defined as:

$$\Phi = {}^1\kappa_\alpha \cdot {}^2\kappa_\alpha / A \quad (13)$$

where A = atom count.

The compressibility of a molecule is a function of the free space between molecules, which must relate to molecular flexibility. Kier and Hall (1999) reported that, for a heterogeneous set of cyclic and acyclic hydrocarbons, compressibility (K_T) correlated well with their Φ values:

$$\begin{aligned} K_T &= 17.785 \Phi + 75.032 \\ n &= 10 \quad r^2 = 0.922 \quad s = 9.4 \end{aligned} \quad (14)$$

Melting point is a function of crystal packing, which also would be expected to relate to molecular flexibility. Eike et al. (2003) obtained a good QSPR for the melting points of 75 quaternary ammonium salts with acyclic saturated alkyl side-chains, using 5 descriptors including Φ ($r^2 = 0.775$).

17 The Variable Connectivity Index

Topological indices are known as graph invariants. However, Randić (1991a, b) introduced the concept of optimization, by using weighted path numbers, of such indices involving heteroatoms in order to obtain better QSAR/QSPR models. The topic then lay dormant until Randić and Basak (1999), Krenkel et al. (2001), Pompe et al. (2004), Randić et al. (2004) and Mu et al. (2009) extended it. Singh et al. (2014) devised some variable Zagreb indices with high discriminating power. Randić and Basak (1999) were able to show that the use of two optimized variable connectivity indices improved QSPR modeling of the boiling points of 58 aliphatic alcohols, with the standard error of prediction being lowered from 6.64° to 3.89°.

Randić (2015) made the impressive point that, in the prediction of boiling points of 100 alcohols, a single variable first-order connectivity index yielded $r^2 = 0.982$, $s = 4.21^\circ$, whereas four non-variable connectivity indices were required to achieve similar statistics ($r^2 = 0.982$, $s = 4.91^\circ$) using the same data.

Randić et al. (2004) have, however, pointed out that as the training set of compounds is changed, the values of the variable indices also change. This means that the method is not fully transferable. It is also likely that unless external test set compounds are very similar to those in the training set, poor external predictivity could result. Additionally there could be a risk that the standard error of prediction could be significantly lower than the experimental error, which is unacceptable (Dearden et al. 2009).

18 Use of Topological Descriptors in Inverse QSAR

It is relatively easy, given a data set of biological activities or physicochemical properties, to describe them quantitatively with a QSAR/QSPR model. But if such a model is to be used predictively, for example to develop more potent drugs, how does one obtain potential drug candidates from the model? The problem is, of course, not restricted to models using topological indices, but Kier and Hall have examined it from that standpoint (Kier et al. 1993a, b; Hall and Kier 1993; Kier and Hall 1993).

Essentially, the approach is (Hall and Kier 1993): (i) set desired target range for property value; (ii) use QSAR equation(s) to obtain target range of each χ index; (iii) convert χ target range into target range of path count; (iv) obtain target range for number of atoms and rings; (v) use interconversion equations to obtain target degree sets; (vi) convert each degree set into a set of corresponding graphs, called candidate graphs; (vii) use best QSAR equation to predict property value for each candidate graph. Kier and Hall (1993) used the above algorithm to design potential isonarcotic agents from a published data set, and found 8 compounds likely to have isonarcotic activity in the desired range.

The Zefirov group has also examined the inverse QSAR problem with the use of a number of topological indices (Baskin et al. 1989; Gordeeva et al. 1990; Zefirov et al. 1991), using an approach quite similar to that of Kier and Hall. Skvortsova et al. (1992, 1993) similarly examined the inverse QSAR problem with the use of kappa indices.

19 Electrotopological State Indices

Almost all molecular descriptors encode essentially either electronic or topological information, and usually represent whole molecules (Hall et al. 1991a). In the early 1990s Kier and Hall developed a set of descriptors that encompass both electronic and topological features, and are atom-based (Kier and Hall 1990; Hall et al. 1991a, b; Hall and Kier 1995); they termed these electrotopological state (E-state) indices. This work was later drawn together in a book (Hall and Kier 1999b).

The electronic factor is considered to relate to the count of non- σ (π and lone-pair) electrons associated with an atom, and is equal to $(\delta^v - \delta)$, where δ^v is the count of valence electrons and δ is the count of σ electrons. The atom intrinsic factor I is defined as $(\delta^v + 1)/\delta$ for first row atoms, and for higher level atoms as $[(2/N)^2 \cdot \delta^v + 1]/\delta$. The perturbation ΔI_i of other atoms j on atom i is defined as:

$$\Delta I_j = \sum_{i=1}^N (I_i - I_j) / r_{ij}^2 \quad (15)$$

where N = principal quantum number, and r = count of atoms in the shortest path connecting atoms i and j , counting both i and j . Hence the E-state value S_i for atom i is $(I_i + \Delta I_i)$.

The power and beauty of the E-state approach are that, unlike most molecular descriptors, it allows an investigator to focus on the effects of individual atoms within a molecule on the activity or property under investigation, and thus potentially aids in determination of mechanism of action. Hall et al. (1991a) found that the ^{17}O NMR chemical shifts for a series of 10 ethers was modelled almost as well by an E-state term ($r^2 = 0.990$, $s = 4.3$) as by a quantum mechanically calculated partial charge ($r^2 = 0.994$, $s = 3.4$).

Another example of E-state correlation with a physicochemical property was given by Hall and Kier (1995), who modeled the boiling points (BP) of 245 alkanes and alcohols:

$$\begin{aligned} \text{BP} &= 8.21 \text{ SsCH}_3 + 14.86 \text{ SssCH}_2 + 24.56 \text{ SsssCH} + 43.76 \text{ SssssC} + 11.63 \text{ SsOH} - 43.95 \\ n &= 245 \quad r^2 = 0.941 \quad s = 8.0 \quad F = 755 \end{aligned} \quad (16)$$

Note that lower case s indicates the number of non-hydrogen bonds formed by each type of atom.

E-state indices were used by Huuskonen et al. (1999) to model the octanol-water partition coefficients ($\log P$) of 300 drugs and related compounds, yielding $r^2 = 0.87$, $s = 0.68$. However, a total of 19 E-state values were required in order to achieve those statistics. Whilst that probably reflected the diversity of the data set, it is not comparable with the results obtained by Abraham et al. (1994), who modeled $\log P$ of 613 diverse organic compounds with only four descriptors, yielding $r^2 = 0.995$, $s = 0.116$.

Kellogg et al. (1996) then introduced E-state values for hydrogen ($I(\text{H})$), mainly to take account of hydrogen bonding. They assumed $I(\text{H})$ to be dependent primarily on the attached atom, and calculated it as $I(\text{H}) = (\delta^v - \delta)^2/\delta$. Rose et al. (2002) found these E-state terms valuable for modeling blood-brain barrier partitioning of 102 diverse compounds, using two hydrogen E-state terms and a χ difference term, with reasonable statistics ($r^2 = 0.66$, $s = 0.45$).

Numerous other workers have found E-state descriptors to be of value in QSAR/QSPR modeling. Ray et al. (2010) used a combination of these and physicochemical descriptors to model the free-radical scavenging activity of 36 hydroxyphenylureas, with $q^2 = 0.957$ and external predictive $r^2 = 0.966$. An exploration of the pharmacophore of some benzodiazepine derivatives as anti-Alzheimer agents was performed by Debnath et al. (2004) using E-state descriptors, with excellent results.

Roy and Mitra (2012) have recently reviewed the use of E-state indices in drug design, property prediction and toxicity assessment. They commented that: "the... use of E-state parameters in the field of computational chemistry portray(s) them as an indispensable tool to expedite investigation of molecular mechanisms and

rational design of molecules, in addition to characterization of physicochemical properties of the molecules and identification of toxic industrial wastes and environmental pollutants”. The present author concurs with those sentiments. In a recent mechanism-based study of compounds causing skin sensitization, Dearden et al. (2015a) found, in 8 QSAR models selected by step-wise regression, that χ values, E-state indices and Kier flexibility terms featured strongly.

20 Biodescriptors

As interest in, and knowledge of, genomics and proteomics increase apace, the biological information available is huge. For example, a proteomics map derived from 2D gel electrophoresis can yield data on the charge, mass and abundance of about 2000 individual proteins (Basak and Gute 2008). The question thus arises as to whether graph theoretical indices can be devised for the characterization of biological data such as DNA sequences or proteomics maps (Basak and Gute 2008). Nandy and Basak (2000) and, Randić et al. (2000) were the first to attempt this. Nandy and Basak examined the effect of toxic substances on DNA primary sequences, and developed simple numerical descriptors from a graphical representation technique that enabled easy visualization of changes in base mutations and deletions arising from toxicity. Nandy et al. (2006) have compared a number of different approaches. The lack of correspondence amongst them led the authors to comment that “until a reasonably dependable characterization system is developed, the underlying graphical systems to be used should be the ones with intuitive appeal to understand the base composition and distribution structure in a sequence, and develop numerical techniques based on such graphs”. Basak and Gute (2008) examined several approaches to the development of mathematical biodescriptors, and concluded that they have a reasonable ability to distinguish between proteomics patterns that result from closely related chemicals and complex mixtures. This could allow the development of new drug candidates, and also act as early warning signals of toxicity. Basak (2010) has reviewed the field up to 2010.

21 Chirality

Graph theoretical indices are 2D descriptors, and so generally cannot distinguish between 3D structural features such as chirality, although Randić et al. (1990) reported an extension of the Randić index approach to give 3D descriptors. A number of attempts have been made to develop TIs that can differentiate between diastereoisomers (Golbraikh et al. 2001) and also between *cis* and *trans* enantiomers (Schultz et al. 1995). Natarajan et al. (2007) have discussed these, and developed a novel approach using a three-point interaction model, whereby the three groups of highest priority attached to a chiral center are viewed from a given

reference point. Attached groups/atoms are assigned δ values by the Kier and Hall method (Kier and Hall 1986), and decreasing importance with increasing topological distance is assigned. The group δ_i value is then calculated as follows:

$$\delta_i^v = \delta_{n1}^v + (\delta_{n2}^v/2) + (\delta_{n3}^v/4) + (\delta_{n4}^v/8) + \dots \quad (16)$$

The relative chirality index (^vRCI) is then calculated as the sum of the δ_i^v values across all relevant bonds in the chiral molecule.

They found that their approach gave good differentiation between diastereoisomers and between enantiomers.

22 Software for Calculation of Topological Indices

There is now a wide range of software available for the calculation of topological indices, and many of these are listed in Table 1. It is sometimes difficult to ascertain whether or not a given software package will calculate a particular topological index, as some websites are not very specific. A number of software websites state, for example, simply that their software will calculate “topological descriptors”, without saying which ones, and such software has not been included in Table 1.

Another matter of potential concern is the accuracy of the topological indices calculated by available software packages, since a number of such programs will probably have been written in-house. A case in point is a paper by Murcia-Soler et al. (2001), who modeled anti-diabetic potencies of drugs using molecular connectivities calculated by their own in-house software. Dearden et al. (2015b) found, in a re-investigation of the Murcia-Soler data, that their reported χ values were incorrect. For example, their $^0\chi^v$ values were all too low by 0.587, and their $^3\chi^v$ values were all too low by 0.230, in comparison with the Molconn-Z values of Hall and Kier, which one would expect to be correct.

23 Conclusions

It is clear from what is written above that there is now a vast literature on the development of topological indices, and on their applications in QSAR and QSPR. One question that has arisen more than once is: are there now enough (or more than enough) topological indices available? It is true that we now have a wide range of TIs available for use in QSAR/QSPR, as this chapter shows. However, one could ask the same question regarding other descriptors, of which we now have thousands (Todeschini and Consonni 2009), yet it does not appear to have been asked, or at least not to the same extent.

Another often-voiced criticism of TIs is that they are difficult to interpret. Kubinyi (1993) implied that he regarded them as “an irrelevance which has had the

unfortunate effect of diverting attention from the real work that needs doing”. Unger (1987), reviewing Kier and Hall’s (1986) book, stated that molecular connectivity “is a highly concocted piece of numerology and is often applied with total lack of rationale”. However, Randić et al. (2016) have made the valid point that although the use of physicochemical properties in QSAR modelling can offer insights into mechanisms of action, they simply relate to parallelisms between such properties and activity, but they tell us nothing about the *structure*-activity relationship directly.

In the present author’s opinion, topological and physicochemical descriptors should be regarded as complementary. Topological indices are one, or perhaps more than one, class of QSAR/QSPR descriptors. Part of the concern about their use is that, on the whole, they have been used as stand-alone descriptors, perhaps even in competition with other types of descriptor, leading to inter-necine rivalry and hence argument: “My descriptors are better than your descriptors”. But why should this be so? All descriptors, of whatever nature and derivation, contain information that could be valuable in modeling, so why not use a descriptor pool of various types of descriptor, as proposed by Basak et al. (1999)? This point has been made strongly by Tseng et al. (2012): “There is no logical reason for keeping descriptor classes segregated. Certainly one can appreciate situations, based upon the endpoint of interest, where multiple classes of descriptors are needed to adequately capture the molecular features and interactions that contribute to the endpoint of interest”. The present author concurs fully with those sentiments. Randić (2008) noted the continuing hostility towards chemical graph theory, and has quoted verbatim many adverse comments by authors and journal editors, although he commented that molecular connectivities have been under less attack of recent years. The present author has witnessed at first hand the verbal abuse of molecular connectivity work in the presence of Kier and Hall at international conferences.

Randić (2008) commented that graph theory is widely appreciated and acknowledged in physics and biology, so why not in chemistry? Who is afraid of graph theory? He also cited many publications in which chemical graph theory has been unfairly attacked, in his view. Randić (2008) put this hostility down to ignorance of the power of graph theory, and speculated as to whether the blindness of critics could “reflect conscious or unconscious concerns how to preserve the monopoly in an applied area of medicinal and physical chemistry”. He expressed similar concerns earlier (Randić 2001) and in a recent co-authored book (Randić et al. 2016). Prelog (1976) has pointed out that “pictorial representations of graphs are so easily intelligible that chemists are often satisfied with inspecting and discussing them without paying too much attention to their algebraic aspects, but it is evident that some familiarity with the theory of graphs is necessary for deeper understanding of their properties”.

It is acknowledged that topological indices are often difficult to interpret in physicochemical terms, although it has to be said that the same applies to a great many non-topological descriptors. A number of authors have attempted such interpretation, and a few such are Basak et al. (1987, 2015), Stankevich et al. (1995), Kier and Hall (2000), Randić et al. (2001), Randić and Zupan (2001) and

Shafiei (2015). Basak (2013a) has also presented a philosophical view of mathematical chemistry.

Although most of the work reported here relates to the application of topological indices in QSAR and QSPR modeling, TIs are also being used in other fields such as proteomics and DNA sequencing (see Sect. 20) and molecular similarity (Basak et al. 1988, 2006; Randić 2014). The latter has potential for database characterization (Cummins et al. 1996) and combinatorial library design (Zheng et al. 1998a, b).

It therefore seems that the future of topological indices and their application to chemistry, biochemistry, biology and medicine are assured for the foreseeable future. There will no doubt continue to be disagreements, but that is the nature of science. If one goes back to 19th century scientific publications, it is clear that bitter arguments were taking place even then.

Let us be grateful that the dire predictions of Auguste Comte (1798–1857) did not come to pass. He wrote: “Every attempt to employ mathematical methods in the study of chemical questions must be considered profoundly irrational and contrary to the spirit of chemistry. If mathematical analysis should ever hold a prominent place in chemistry—an aberration which is happily almost impossible—it would occasion a rapid and widespread degeneration of that science” (Liang et al. 1993).

References

- Abraham, M. H., Chadha, H. S., & Mitchell, R. C. (1994). Hydrogen bonding. Part 32. An analysis of water-octanol and water-cyclohexane partitioning and the log P parameter of Seiler. *Journal of Pharmaceutical Sciences*, 83, 1085–1100.
- Abreu, R. M. V., Ferreira, I. C. F. R., & Queiroz, M. J. R. P. (2009). QSAR model for predicting radical scavenging activity of di(hetero)arylamines derivatives of benzo[b]thiophenes. *European Journal of Medicinal Chemistry*, 44, 1952–1958.
- Aptula, A. O., Jeliaskova, N. G., Schultz, T. W., & Cronin, M. T. D. (2005). The better predictive model: high q^2 for the training set or low root mean square error of prediction for the test set? *QSAR & Combinatorial Science*, 24, 385–396.
- Bajaj, S., Sami, S. S., & Madan, A. K. (2005). Prediction of anti-inflammatory activity of *N*-arylanthranilic acids: Computational approach using refined Zagreb indices. *Croatica Chemica Acta*, 78, 165–174.
- Balaban, A. T. (1979). Chemical graphs. 34. Five new topological indices for the branching of tree-like graphs. *Theoretica Chimica Acta*, 53, 355–375.
- Balaban, A. T. (1982). Highly discriminating distance-based topological index. *Chemical Physics Letters*, 89, 399–404.
- Balaban, A. T. (1985). Applications of graph theory in chemistry. *Journal of Chemical Information and Computer Sciences*, 25, 334–343.
- Basak, S. C. (1987). Use of molecular complexity indices in predictive pharmacology and toxicology: A QSAR approach. *Medical Science Research*, 15, 605–609.
- Basak, S. C., Magnuson, V. R., Niemi, G. J., Regal, R. R., & Veith, G. D. (1987). Topological indices: Their nature, mutual relatedness, and applications. *Mathematical Modelling*, 8, 300–305.
- Basak, S. C., Magnuson, V. R., Niemi, G. J., & Regal, R. R. (1988). Determining structural similarity of chemical using graph-theoretic indices. *Discrete Applied Mathematics*, 19, 17–44.

- Basak, S. C. (1999). Information theoretic indices of neighborhood complexity and their applications. In J. Devillers & A. T. Balaban (Eds.), *Topological indices and related descriptors in QSAR and QSPR* (pp. 563–593). Amsterdam: Gordon and Breach Science Publishers.
- Basak, S. C., Gute, B. D., & Grunwald, G. D. (1999). A hierarchical approach to the development of QSAR models using topological, geometrical and quantum chemical parameters. In J. Devillers & A. T. Balaban (Eds.), *Topological indices and related descriptors in QSAR and QSPR* (pp. 563–593). Amsterdam: Gordon and Breach Science Publishers.
- Basak, S. C., Balaban, A. T., Grunwald, G. D., & Gute, B. D. (2000). Topological indices: Their nature and mutual relatedness. *Journal of Chemical Information and Computer Sciences*, 40, 891–898.
- Basak, S. C., Gute, B. D., & Mills, D. (2006). Similarity methods in analog selection, property estimation and clustering of diverse chemicals. *ARKIVOC*, 157–210.
- Basak, S. C., & Gute, B. D. (2008). Mathematical biodescriptors of proteomics maps: Background and applications. *Current Opinion in Drug Discovery & Development*, 11, 320–326.
- Basak, S. C. (2010). Role of mathematical chemodescriptors and proteomics-based biodescriptors in drug discovery. *Drug Development Research*, 72, 1–9.
- Basak, S. C. (2013a). Philosophy of mathematical chemistry: A personal perspective. *HYLE—International Journal for Philosophy of Chemistry*, 19, 4–17.
- Basak, S. C. (2013b). Mathematical descriptors for the prediction of property, bioactivity, and toxicity of chemicals from their structure: A chemical-cum-biochemical approach. *Current Computer-Aided Drug Design*, 9, 449–462.
- Basak, S. C., Grunwald, G. D., & Majumdar, S. (2015). Intrinsic dimensionality of chemical space: Characterization and applications. *Mol2Net Section B, Proceedings*, 1, 1–10.
- Baskin, I. I., Gordeeva, E. V., Devdarian, R. O., Zefirov, N. S., Palyulin, V. A., & Stankevich, M. I. (1989). Methodology of solution of the inverse problem for the structure-property relationship for the case of topological indices. *Doklady Chemistry*, 307, 217–220.
- Bharate, S. B., & Singh, I. P. (2011). Quantitative structure-activity relationship study of phloroglucinol-terpene adducts as anti-leishmanial agents. *Bioorganic & Medicinal Chemistry Letters*, 21, 4310–4315.
- Burkhard, L. P., Andren, A. W., & Armstrong, D. E. (1983). Structure activity relationships using molecular connectivity indices with principal component analysis. *Chemosphere*, 12, 935–943.
- Caballero, J., Fernández, M., Saavedra, M., & González-Nilo, F. D. (2008). 2D autocorrelation, CoMFA, and CoMSIA modeling of protein tyrosine kinases' inhibition by substituted pyrido [2,3-*d*]pyrimidine derivatives. *Bioorganic & Medicinal Chemistry*, 16, 810–821.
- Cronin, M. T. D., et al. (2002). Comparative assessment of methods to develop QSARs for the prediction of the toxicity of phenols to *Tetrahymena pyriformis*. *Chemosphere*, 49, 1201–1221.
- Cummins, D. J., Andrews, C. W., Bentley, J. A., & Cory, M. (1996). Molecular diversity and chemical databases: Comparison of medicinal chemistry knowledge bases and databases of commercially available compounds. *Journal of Chemical Information and Computer Sciences*, 36, 750–763.
- Dearden, J. C., Bradburne, S. J. A., Cronin, M. T. D., Solanki, P., et al. (1988). The physical significance of molecular connectivity. In J. E. Turner, M. W. England, T. W. Schultz, & N. J. Kwaak (Eds.), *QSAR-88* (pp. 43–50). Oakridge, TN: U.S. Department of Energy.
- Dearden, J. C., Cronin, M. T. D., Schultz, T. W., & Lin, D. T. (1995). QSAR study of the toxicity of nitrobenzenes to *Tetrahymena pyriformis*. *Quantitative Structure-Activity Relationships*, 14, 189–196.
- Dearden, J. C., Wong, E. H. Y., & Walker, J. D. (2004). Molecular connectivity: polarity correction to improve correlation with hydrophobicity. *QSAR & Combinatorial Science*, 23, 75–79.
- Dearden, J. C., Cronin, M. T. D., & Kaiser, K. L. E. (2009). How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR). *SAR and QSAR in Environmental Research*, 20, 241266.
- Dearden, J. C., et al. (2015a). Mechanism-based QSAR modeling of skin sensitization. *Chemical Research in Toxicology*, 28, 1975–1968.

- Dearden, J. C., Hewitt, M., & Rowe, P. H. (2015b). QSAR study of some anti-hyperglycaemic sulphonylurea drugs. *SAR & QSAR in Environmental Research*, 26, 439–448.
- Debnath, B., Gayen, S., Basu, A., Srikanth, K., & Jha, T. (2004). Quantitative structure-activity relationship study on some benzodiazepine derivatives as anti-Alzheimer agents. *Journal of Molecular Modelling*, 10, 328–334.
- Devillers, J. (1999a). No-free-lunch molecular descriptors in QSAR and QSPR. In J. Devillers & A. T. Balaban (Eds.), *Topological indices and related descriptors in QSAR and QSPR* (pp. 1–20). Amsterdam: Gordon and Breach Science Publishers.
- Devillers, J. (1999b). Autocorrelation descriptors for modeling (eco)toxicological endpoints. In J. Devillers & A. T. Balaban (Eds.), *Topological indices and related descriptors in QSAR and QSPR* (pp. 595–612). Amsterdam: Gordon and Breach Science Publishers.
- Dureja, H., Gupta, S., & Madan, A. K. (2008). Topological models for prediction of pharmacokinetic parameters of cephalosporins using random forest, decision tree and moving average analysis. *ScientiaPharmaceutica*, 76, 377–394.
- Eike, D. M., Brennecke, J. F., & Maginn, E. J. (2003). Predicting melting points of quaternary ammonium ionic liquids. *Green Chemistry*, 5, 323–328.
- Estrada, E. (2002). Physicochemical interpretation of molecular connectivity indices. *Journal of Physical Chemistry A*, 106, 9085–9091.
- Filip, P. A., Balaban, T. S., & Balaban, A. T. (1987). A new approach for devising graph invariants: Derived topological indices with low degeneracy and good correlation ability. *Journal of Mathematical Chemistry*, 1, 61–83.
- Golbraikh, A., & Tropsha, A. (2002). Beware of q^2 ! *Journal of Molecular Graphics and Modelling*, 20, 269–276.
- Golbraikh, A., Bonchev, D., & Tropsha, A. (2001). Novel chirality descriptors derived from molecular topology. *Journal of Chemical Information and Computer Sciences*, 41, 147–158.
- González, M. P., Caballero, J., Helguera, A. M., et al. (2006). 2D autocorrelation modelling of the inhibitory activity of cytokinin-derived cyclin-dependent kinase inhibitors. *Bulletin of Mathematical Biology*, 68, 735–751.
- Gordeeva, E. V., Molchanova, M. S., & Zefirov, N. S. (1990). General methodology and computer program for the exhaustive restoring of chemical structures by molecular connectivity indexes. *Tetrahedron Computer Methodology*, 3, 389–415.
- Gordon, M., & Scantlebury, G. R. (1964). Non-random polycondensation: statistical theory of the substitution effect. *Transactions of the Faraday Society*, 60, 604–621.
- Gramatica, P., Pilutti, P., & Papa, E. (2007). Approaches for externally validated QSAR modelling of nitrated polycyclic aromatic hydrocarbon mutagenicity. *SAR and QSAR in Environmental Research*, 18, 169–178.
- Gupta, M. K., Sagar, R., Shaw, A. K., & Prabhakar, Y. S. (2005). CP-MLR directed QSAR studies on the antimycobacterial activity of functionalized alkenols—Topological descriptors in modeling the activity. *Bioorganic & Medicinal Chemistry*, 13, 35–343.
- Gute, B. D., & Basak, S. C. (1997). Predicting acute toxicity (LC_{50}) of benzene derivatives using theoretical molecular descriptors: A hierarchical QSAR approach. *SAR & QSAR in Environmental Research*, 7, 117–131.
- Gutman, I., & Trinajstić, N. (1972). Graph theory and molecular orbitals. Total π -electron energy of alternant hydrocarbons. *Chemical Physics Letters*, 17, 535–538.
- Gutman, I., & Körtvélyesi, T. (1995). Wiener indices and molecular surfaces. *Zeitschrift für Naturforschung*, 50A, 669–671.
- Hall, L. H., Kier, L. B., & Murray, W. J. (1975). Molecular connectivity II: Relationship to water solubility and boiling point. *Journal of Pharmaceutical Sciences*, 64, 1974–1977.
- Hall, L. H., Maynard, E. L., & Kier, L. B. (1989). QSAR investigation of benzene toxicity to fathead minnow using molecular connectivity. *Environmental Toxicology and Chemistry*, 8, 783–788.
- Hall, L. H., Mohney, B., & Kier, L. B. (1991a). The electrotopological state: Structure information at the atomic level for molecular graphs. *Journal of Chemical Information and Computer Sciences*, 31, 76–82.

- Hall, L. H., Mohney, B., & Kier, L. B. (1991b). The electrotopological state: An atom index for QSAR. *Quantitative Structure-Activity Relationships*, 10, 43–51.
- Hall, L. H., & Kier, L. B. (1993). Design of molecules from quantitative structure-activity relationship models. III. Role of higher order path counts: Path three. *Journal of Chemical Information and Computer Sciences*, 33, 598–603.
- Hall, L. H., & Kier, L. B. (1995). Electrotopological state indices for atom types: A novel combination of electronic, topological, and valence state information. *Journal of Chemical Information and Computer Sciences*, 35, 1039–1045.
- Hall, L. H., & Kier, L. B. (1999a). Molecular connectivity chi indices for database analysis and structure-property modeling. In J. Devillers & A. T. Balaban (Eds.), *Topological indices and related descriptors in QSAR and QSPR* (pp. 307–360). Amsterdam: Gordon and Breach Science Publishers.
- Hall, L. H., & Kier, L. B. (1999b). *Molecular structure description: The electrotopological state*. San Diego: Academic Press.
- Hansch, C., Gao, H., & Hoekman, D. (1998). A generalized approach to comparative QSAR. In J. Devillers (Ed.), *Comparative QSAR* (pp. 285–368). London: Taylor & Francis.
- Hollas, B., Gutman, I., & Trinajstić, N. (2005). On reducing correlations between topological indices. *Croatica Chemica Acta*, 78, 489–492.
- Hosoya, H. (1971). Topological index. A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. *Bulletin of the Chemical Society of Japan*, 44, 2332–2339.
- Huuskonen, J. J., Villa, A. E. P., & Tetko, I. V. (1999). Prediction of partition coefficient based on atom-type electrotopological state indices. *Journal of Pharmaceutical Sciences*, 88, 229–233.
- Ivanciuc, O. (2000). QSAR comparative study of Wiener descriptors for weighted molecular graphs. *Journal of Chemical Information and Computer Sciences*, 40, 1412–1422.
- Ivanciuc, O., & Balaban, A. T. (1999). The graph description of chemical structures. In J. Devillers & A. T. Balaban (Eds.), *Topological indices and related descriptors in QSAR and QSPR* (pp. 59–167). Amsterdam: Gordon and Breach Science Publishers.
- Iwamura, H. (1980). Structure-taste relationship of perillartine and nitro- and cyanoaniline derivatives. *Journal of Medicinal Chemistry*, 23, 308–312.
- Jalali-Heravi, M., & Asadollahi-Baboli, M. (2008). QSAR analysis of platelet-derived growth inhibitors using GA-ANN and shuffling crossvalidation. *QSAR & Combinatorial Science*, 27, 750–757.
- Johnson, S. R. (2008). The trouble with QSAR (or how I learned to stop worrying and embrace fallacy). *Journal of Chemical Information and Modeling*, 48, 25–26.
- Kellogg, G. E., Kier, L. B., Gaillard, P., & Hall, L. H. (1996). The E-state fields: Applications to 3D QSAR. *Journal of Computer-Aided Molecular Design*, 10, 513–520.
- Kier, L. B., & Hall, L. H. (1976a). Molecular connectivity VII: Specific treatment of heteroatoms. *Journal of Pharmaceutical Sciences*, 65, 1806–1809.
- Kier, L. B., & Hall, L. H. (1976b). *Molecular connectivity in chemistry and drug research*. New York: Academic Press.
- Kier, L. B., Di Paolo, T., & Hall, L. H. (1977). Structure-activity studies on odor molecules using molecular connectivity. *Journal of Theoretical Biology*, 67, 585–595.
- Kier, L. B., & Hall, L. H. (1978). Molecular connectivity study of muscarinic receptor affinity of acetylcholine antagonists. *Journal of Pharmaceutical Sciences*, 67, 1408–1412.
- Kier, L. B. (1985). A shape index from molecular graphs. *Quantitative Structure-Activity Relationships*, 4, 109–116.
- Kier, L. B., & Hall, L. H. (1986). *Molecular connectivity in structure-activity analysis*. Chichester: Research Studies Press.
- Kier, L. B. (1986a). Shape indexes of orders one and three from molecular graphs. *Quantitative Structure-Activity Relationships*, 5, 1–7.
- Kier, L. B. (1986b). Distinguishing atom differences in a molecular graph shape index. *Quantitative Structure-Activity Relationships*, 5, 7–12.

- Kier, L. B. (1989). An index of molecular flexibility from kappa shape attributes. *Quantitative Structure-Activity Relationships*, 8, 221–224.
- Kier, L. B., & Hall, L. H. (1990). An electrotopological-state index for atoms in molecules. *Pharmaceutical Research*, 7, 801–807.
- Kier, L. B., & Hall, L. H. (1991). A differential molecular connectivity index. *Quantitative Structure-Activity Relationships*, 10, 134–140.
- Kier, L. B., & Hall, L. H. (1993). The generation of molecular structures from a graph-based QSAR equation. *Quantitative Structure-Activity Relationships*, 12, 383–388.
- Kier, L. B., Hall, L. H., & Frazer, J. W. (1993a). Design of molecules from quantitative structure-activity relationship models. I. Information transfer between path and vertex degree counts. *Journal of Chemical Information and Computer Sciences*, 33, 143–147.
- Kier, L. B., Hall, L. H., & Frazer, J. W. (1993b). Design of molecules from quantitative structure-activity relationship models. II. Derivation and proof of information transfer relating equations. *Journal of Chemical Information and Computer Sciences*, 33, 148–152.
- Kier, L. B., & Hall, L. H. (1999). The kappa indices for modeling molecular shape and flexibility. In J. Devillers & A. T. Balaban (Eds.), *Topological indices and related descriptors in QSAR and QSPR* (pp. 455–489). Amsterdam: Gordon and Breach Science Publishers.
- Kier, L. B., & Hall, L. H. (2000). Intermolecular accessibility: The meaning of molecular connectivity. *Journal of Chemical Information and Computer Sciences*, 40, 792–795.
- Kier, L. B., & Hall, L. H. (2001). Molecular connectivity: Intermolecular accessibility and encounter simulation. *Journal of Molecular Graphics and Modelling*, 20, 76–83.
- Kier, L. B., Hall, L. H., Murray, W. J., & Randić, M. (1975). Molecular connectivity I: Relationship to local anesthesia. *Journal of Pharmaceutical Sciences*, 64, 1971–1974.
- Kier, L. B., Murray, W. J., Randić, M., & Hall, L. H. (1976). Molecular connectivity V: Connectivity series concept applied to density. *Journal of Pharmaceutical Sciences*, 65, 1226–1230.
- Krenkel, G., Castro, E. A., & Toropov, A. A. (2001). The variable molecular descriptors based on the optimization of correlation weights of local graph invariants. *Journal of Molecular Structure (THEOCHEM)*, 542, 107–113.
- Krishnasamy, C., Raghuraman, A., Kier, L. B., & Desai, U. R. (2008). Application of molecular connectivity and electro-topological indices in quantitative structure-activity analysis of pyrazole derivatives as inhibitors of factor Xa and thrombin. *Chemistry & Biodiversity*, 5, 2609–2620.
- Kubinyi, H. (1993). *QSAR: Hansch analysis and related approaches*. Weinheim: Wiley-VCH.
- Kubinyi, H. (1997). A general view on similarity and QSAR studies. In H. van de Waterbeemd, B. Testa, & G. Folkers (Eds.), *Computer-assisted lead finding and optimization: Current tools for medicinal chemistry* (pp. 9–28). Basel, Verlag Helvetica Chimica Acta, & Weinheim, VCH.
- Li, X. H., Jalbout, A. F., & Solimannejad, M. (2003). Definition and application of a novel valence molecular connectivity index. *Journal of Molecular Structure (THEOCHEM)*, 663, 81–85.
- Liang, Y.-Z., Kvalheim, O. M., & Manne, R. (1993). White, grey and black multicomponent systems: A classification of mixture problems and methods for their quantitative analysis. *Chemometrics and Intelligent Laboratory Systems*, 18, 235–250.
- Livingstone, D. (1995). *Data analysis for chemists: Applications to QSAR and chemical product design*. Oxford: Oxford University Press.
- Livingstone, D. J. (2000). The characterization of chemical structures using molecular properties. A survey. *Journal of Chemical Information and Computer Sciences*, 40, 195–209.
- Lopez de Compadre, R. L., Compadre, C. M., Castillo, R., & Dunn, W. J. (1983). On the use of connectivity indices in quantitative structure activity studies. *European Journal of Medicinal Chemistry*, 18, 569–571.
- Lu, X., Tao, S., Cao, J., & Dawson, R. W. (1999). Prediction of fish bioconcentration factors of nonpolar organic pollutants based on molecular connectivity indices. *Chemosphere*, 39, 987–999.
- Luisi, P. L. (1977). Molecular conformational rigidity: An approach to quantification. *Naturwissenschaften*, 64, 569–574.

- Maran, U., Sild, S., Tulp, I., Takkis, K., & Moosus, M. (2010). Molecular descriptors from two-dimensional chemical structure. In M. T. D. Cronin & J. C. Madden (Eds.), *In silico toxicology: Principles and applications* (pp. 148–192). Cambridge: RSC Publishing.
- Mekenyan, O., Bonchev, D., Sabljic, A., & Trinajstić, N. (1987). Application of topological indices to QSAR. The use of the Balaban index and the electrotopology index for correlations with toxicity of ethers on mice. *Acta Pharmaceutica Jugoslavica*, 37, 75–86.
- Moreau, G., & Broto, P. (1980). The autocorrelation of a topological structure: A new molecular descriptor. *Nouveau Journal de Chimie*, 4, 359–360.
- Mowshowitz, A. (1968). Entropy and the complexity of graphs: I. An index of the relative complexity of a graph. *Bulletin of Mathematical Biophysics*, 30, 175–204.
- Mu, L., He, H., Yang, W., & Feng, C. (2009). Variable molecular connectivity indices for predicting the diamagnetic susceptibilities of organic compounds. *Industrial and Engineering Chemistry Research*, 48, 4165–4175.
- Murcia-Soler, M., Pérez-Giménez, F., Nalda-Molina, R., Salabert-Salvador, M. T., Garcia-March, F. J., Cercós-del-Pozo, R. A., et al. (2001). QSAR analysis of hypoglycaemic agents using the topological indices. *Journal of Chemical Information and Computer Sciences*, 41, 1345–1354.
- Murray, W. J., Hall, L. H., & Kier, L. B. (1975). Molecular connectivity III. Relationship to partition coefficients. *Journal of Pharmaceutical Sciences*, 64, 1978–1981.
- Nandy, A., & Basak, S. C. (2000). Simple numerical descriptor for quantifying effect of toxic substances on DNA sequences. *Journal of Chemical Information and Computer Sciences*, 40, 915–919.
- Nandy, A., Harle, M., & Basak, S. C. (2006). Mathematical descriptors of DNA sequences: Development and applications. *ARKIVOC*, 1x, 211–238.
- Natarajan, R., Basak, S. C., & Neumann, T. S. (2007). Novel approach for the numerical characterization of molecular chirality. *Journal of Chemical Information and Modeling*, 47, 771–775.
- Netzeva, T. I. (2004). Whole molecule and atom-based topological descriptors. In M. T. D. Cronin & D. J. Livingstone (Eds.), *Predicting chemical toxicity and fate* (pp. 61–83). Boca Raton: FL, CRC Press.
- Nikolić, S., Kovačević, G., Miličević, A., & Trinajstić, N. (2003). The Zagreb indices 30 years after. *Croatica Chemica Acta*, 76, 113–124.
- Noorizadeh, H., Farmany, A., & Noorizadeh, M. (2011). Quantitative structure-retention relationships analysis of retention index of essential oils. *Quimica Nova*, 34, 242–249.
- OECD (2004). Validation of (Q)SAR models. Retrieved May 26, 2016 from <http://www.oecd.org/chemicalsafety/risk-assessment/validationofqsarmodels.htm>.
- Pal, D. K., Sengupta, C., & De, A. U. (1988). A new topochemical descriptor (TAU) in molecular connectivity concept: Part I—Aliphatic compounds. *Indian Journal of Chemistry*, 27B, 734–739.
- Platt, J. R. (1947). Influence of neighbor bonds on additive bond properties in paraffins. *Journal of Chemical Physics*, 15, 419–420.
- Pompe, M., Veber, M., Randić, M., & Balaban, A. T. (2004). Using variable and fixed topological indices for the prediction of reaction rate constants of volatile unsaturated hydrocarbons with OH radicals. *Molecules*, 9, 1160–1176.
- Prelog, V. (1976). Foreword. In A. T. Balaban (Ed.), *Chemical applications of graph theory*. London: Academic Press.
- Randić, M. (1975). On characterization of molecular branching. *Journal of the American Chemical Society*, 97, 6609–6615.
- Randić, M., Jerman-Blažić, B., & Trinajstić, N. (1990). Development of 3-dimensional molecular descriptors. *Computers & Chemistry*, 14, 237–246.
- Randić, M. (1991a). Novel graph theoretical approach to heteroatoms in quantitative structure-activity relationships. *Chemometrics and Intelligent Laboratory Systems*, 10, 213–227.
- Randić, M. (1991b). On computation of optimal parameters for multivariate analysis of structure-property relationship. *Journal of Computational Chemistry*, 12, 970–980.

- Randić, M., & Basak, S. C. (1999). Optimal molecular descriptors based on weighted path numbers. *Journal of Chemical Information and Computer Sciences*, 39, 261–266.
- Randić, M., Vracko, M., Nandy, A., & Basak, S. C. (2000). On 3D graphical representation of DNA primary sequences and their numerical characterization. *Journal of Chemical Information and Computer Sciences*, 40, 1235–1244.
- Randić, M. (2001). The connectivity index 25 years after. *Journal of Molecular Graphics and Modelling*, 20, 19–35.
- Randić, M., & Zupan, J. (2001). On interpretation of well-known topological indices. *Journal of Chemical Information and Computer Sciences*, 41, 550–560.
- Randić, M., Balaban, A. T., & Basak, S. C. (2001). On structural interpretation of several distance related topological indices. *Journal of Chemical Information and Computer Sciences*, 41, 593–601.
- Randić, M., Pompe, M., Mills, D., & Basak, S. C. (2004). Variable connectivity index as a tool for modeling structure-property relationships. *Molecules*, 9, 1177–1193.
- Randić, M. (2008). On history of the Randić index and emerging hostility toward chemical graph theory. *MATCH Communications in Mathematical and in Computer Chemistry*, 59, 5–124.
- Randić, M. (2014). On of molecular similarity based on a single descriptor. *Chemical Physics Letters*, 599, 1–6.
- Randić, M. (2015). On the history of the connectivity index: From the connectivity index to the exact solution of the protein alignment problem. *SAR & QSAR in Environmental Research*, 26, 523–555.
- Randić, M., Novič, M., & Plavšić, D. (2016). *Solved and unsolved problems of structural chemistry*. Boca Raton, FL: CRC Press.
- Ray, S., Roy, P. P., Sengupta, C., & Roy, K. (2010). Exploring QSAR of hydroxyphenylureas as antioxidants using physicochemical and electrotopological state atom parameters. *Molecular Simulation*, 36, 484–492.
- Ribo, J. M., & Kaiser, K. L. E. (1984). Toxicities of chloroanilines to *Photobacterium phosphoreum* and their correlations and effects on other organisms and structural parameters. In K. L. E. Kaiser (Ed.), *QSAR in environmental toxicology* (pp. 319–336). Dordrecht: D. Reidel Publishing Co.
- Rose, K., Hall, L. H., & Kier, L. B. (2002). Modeling blood-brain barrier partitioning using the electrotopological state. *Journal of Chemical Information and Computer Sciences*, 42, 651–666.
- Rouvray, D. H., & Crafford, B. C. (1976). The dependence of physicochemical properties on topological factors. *South African Journal of Science*, 72, 47–51.
- Rouvray, D. H., & King, R. B. (Eds.) (2002). *Topology in chemistry*. Chichester: Horwood Publishing Limited.
- Roy, K., & Ghosh, G. (2003). Introduction of extended topochemical atom (ETA) indices in the valence electron mobile (VEM) environment as tools for QSAR/QSPR studies. *Internet Electronic Journal of Molecular Design*, 2, 599–620.
- Roy, K., & Ghosh, G. (2004). QSTR with extended topochemical atom indices. 2. Fish toxicity of substituted benzenes. *Journal of Chemical Information & Computer Sciences*, 44, 559–567.
- Roy, K., & Ghosh, G. (2009). QSTR with extended topochemical atom (ETA) indices. 12. QSAR for the toxicity of diverse aromatic compounds to *Tetrahymena pyriformis*. *Chemosphere*, 77, 999–1009.
- Roy, K., & Mitra, I. (2012). Electrotopological state atom (E-state) index in drug design, QSAR, property prediction and toxicity assessment. *Current Computer-Aided Drug Design*, 8, 135–158.
- Roy, K., Kar, S., & Das, R. N. (2015). *Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment*. Amsterdam: Academic Press.
- Roy, K., & Saha, A. (2003). QSPR with TAU indices: Water solubility of diverse functional acyclic compounds. *Internet Electronic Journal of Molecular Design*, 2, 475–491. http://biochempress.com/Files/iejmd_2003_2_0475.pdf.

- Sabljić, A. (1985). Calculation of retention indices by molecular topology: Chlorinated benzenes. *Journal of Chromatography*, 319, 1–8.
- Sabljić, A. (1990). Topological indices and environmental chemistry. In W. Karcher & J. Devillers (Eds.), *Practical applications of quantitative structure-activity relationships (QSAR) in environmental chemistry and toxicology* (pp. 61–82). Dordrecht: Kluwer Academic Publishers.
- Sabljić, A., & Protić-Sabljić, M. (1983). Quantitative structure-activity study of the mechanism of inhibition of microsomal p-hydroxylation of aniline by alcohols. Role of steric factors. *Molecular Pharmacology*, 23, 213–218.
- Schultz, H. P., Schultz, E. B., & Schultz, T. P. (1995). Topological organic chemistry. 9. Graph theory and molecular topological indices of stereoisomeric organic compounds. *Journal of Chemical Information & Computer Sciences*, 35, 864–870.
- Shafiei, F. (2015). Relationship between topological indices and thermodynamic properties and of the monocarboxylic acids applications in QSPR. *Iranian Journal of Mathematical Chemistry*, 6, 15–28.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423.
- Singh, M., Das, K. C., Gupta, S., & Madan, A. K. (2014). Refined variable Zagreb indices: Highly discriminating topological descriptors for QSAR/QSPR. *International Journal of Chemical Modeling*, 6, 403–428.
- Skvortsova, M. I., Baskin, I. I., Slovokhotova, O. L., Palyulin, V. A., & Zefirov, N. S. (1992). Inverse problem in QSAR/QSPR analysis for case of topological indexes characterizing molecular form (Kier indexes). *Doklady Chemistry*, 324, 103–107.
- Skvortsova, M. I., Baskin, I. I., Slovokhotova, O. L., Palyulin, V. A., & Zefirov, N. S. (1993). Inverse problem in QSAR/QSPR studies for the case of topological indices characterizing molecular shape (Kier indices). *Journal of Chemical Information & Computer Sciences*, 33, 630–634.
- Solomon, K. A., Sundararajan, S., & Abirami, V. (2009). QSAR studies on *N*-aryl derivative activity towards Alzheimer's disease. *Molecules*, 14, 1448–1455.
- Stankevich, I. V., Skvortsova, M. I., & Zefirov, N. S. (1995). On a quantum chemical interpretation of molecular connectivity indices for conjugated hydrocarbons. *Journal of Molecular Structure (THEOCHEM)*, 342, 173–179.
- Stiel, L. I., & Thodos, G. (1962). The normal boiling points and critical constants of saturated aliphatic hydrocarbons. *American Institute of Chemical Engineers Journal*, 8, 527–529.
- Thakur, A., Thakur, M., Khadikar, P. V., Supuran, C. T., & Sudele, P. (2004). QSAR study on benzenesulphonamide carbonic anhydrase inhibitors: Topological approach using Balaban index. *Bioorganic & Medicinal Chemistry*, 12, 789–793.
- Todeschini, R., & Gramatica, P. (1997). The WHIM theory: New 3D molecular descriptors for QSAR in environmental modelling. *SAR and QSAR in Environmental Research*, 7, 89–115.
- Todeschini, R., & Consonni, V. (2009). *Molecular descriptors for chemoinformatics. Alphabetical listing* (2nd ed., Vol. 1). Weinheim: Wiley-VCH.
- Todeschini, R., Lasagni, M., & Marengo, E. (1994). New molecular descriptors for 2D and 3D structures theory. *Journal of Chemometrics*, 8, 263–272.
- Tong, J., Liu, S., Zhou, P., Wu, B., & Li, Z. (2008). A novel descriptor of amino acids and its application in peptide QSAR. *Journal of Theoretical Biology*, 253, 90–97.
- Tseng, Y. J., Hopfinger, A. J., & Esposito, E. X. (2012). The great descriptor melting pot: Mixing descriptors for the common good of QSAR models. *Journal of Computer-Aided Molecular Design*, 26, 39–43.
- Unger, S. H. (1987). Molecular connectivity in structure-activity analysis (book review). *Journal of Pharmaceutical Sciences*, 76, 269–270.
- Vlaia, V., Olariu, T., Vlaia, L., Butur, M., Ciubotariu, C., Medeleanu, M., et al. (2009). Quantitative structure-activity relationship (QSAR). IV. Analysis of the toxicity of aliphatic esters by means of weighted holistic invariant molecular (WHIM) descriptors. *Farmacia*, 57, 511–522.
- Wells, P. R. (1968). *Linear free energy relationships*. London: Academic Press.

- Wiener, H. (1947). Structural determination of paraffin boiling points. *Journal of the American Chemical Society*, 69, 17–20.
- Wiener, H. (1948). Relation of the physical properties of the isomeric alkanes to molecular structure. *Journal of Physical Chemistry*, 52, 1082–1089.
- Wilson, E. B. (1952). *Introduction to scientific research*. New York: McGraw-Hill.
- Yilmaz, B., & Göktürk, M. (2009). Interactive data mining for molecular graphs. *Journal of Automated Methods & Management in Chemistry*. doi:[10.1155/2009/502527](https://doi.org/10.1155/2009/502527).
- Zefirov, N. S., Palyulin, V. A., & Radchenko, E. V. (1991). Problem of generation of structures with definite properties. Solution of inverse problem for Balaban centric index. *Doklady Chemistry*, 316, 921–924.
- Zhang, S. G., Lei, W., Xia, M. Z., & Wang, F. W. (2005). QSAR study of N-containing corrosion inhibitors: Quantum chemical approach assisted by topological index. *Journal of Molecular Structure (THEOCHEM)*, 732, 173–182.
- Zheng, W., Cho, S. J., & Tropsha, A. (1998a). Rational combinatorial design. 1. Focus 2-D: A new approach to the design of targeted combinatorial chemical libraries. *Journal of Chemical Information and Computer Sciences*, 38, 251–258.
- Zheng, W., Cho, S. J., & Tropsha, A. (1998b). Rational combinatorial design. 2. Rational design of targeted combinatorial peptide libraries using chemical similarity probe and inverse QSAR approaches. *Journal of Chemical Information and Computer Sciences*, 38, 259–268.

Advances in QSAR Modeling

Applications in Pharmaceutical, Chemical, Food,
Agricultural and Environmental Sciences

Roy, K. (Ed.)

2017, X, 555 p. 132 illus., 71 illus. in color., Hardcover

ISBN: 978-3-319-56849-2