J Math Chem (2013) 51:811–816
DOI 10.1007/s10910-013-0143-x

BRIEF COMMUNICATION

A simple electrotopological index for quantitative structure-activity relationship correlation of physical properties with biomolecular activities

Ian N. Weaver · Donald F. Weaver

Received: 21 August 2012 / Accepted: 3 January 2013 / Published online: 13 January 2013

Abstract Quantitative structure-activity relationship (QSAR) studies constitute a process by which the physicochemical properties of a set of chemical structures are quantitatively correlated with a measurable, such as the concentration of a substance required to give a certain therapeutic drug response. 2D-QSAR studies start with 10–20 analogues, ranging from biologically active to inactive; each analogue, regardless of bioactivity, is described by a series of descriptors. To further broaden the practical utility of these simple descriptors we have sought to identify hybrid indices which are trivial to calculate but which capture data from at least two categories of descriptors. An electrotopological descriptor, termed ET_Z, which combines electronic information and molecular topology, has been devised and validated against a set of 25 anticonvulsant hydantoin molecules. This ET_Z is based solely on atomic connectivity information obtained from the graph without explicit input from molecular geometry.

Keywords Quantitative structure-activity relationship study · Descriptors · Topological index · Electrotopological index · Computer-aided drug design

Quantitative structure-activity relationship (QSAR) studies constitute a process by which the physicochemical properties of a defined set of chemical structures are

I. N. Weaver · D. F. Weaver (✉)
Departments of Biomedical Engineering, Dalhousie University, Halifax, NS, Canada
e-mail: donald.weaver@dal.ca

D. F. Weaver
Departments of Chemistry, Dalhousie University, Halifax, NS, Canada

D. F. Weaver
Departments of Medicine (Neurology), Dalhousie University, Halifax, NS, Canada

quantitatively correlated with a rigorously defined measurable, such as the concentration of a substance required to give a certain biological response. QSAR is widely used during the optimization phase of drug design [1–7].

Over the past 30 years, QSAR has progressed from the simple regression equations of Hansch (1D-QSAR) to more modern 2D-QSAR [8–10]. Typically, 2D-QSAR studies start with 10–20 analogues, ranging from biologically active to inactive. Each analogue, regardless of bioactivity, is described by a series of descriptors which can be divided into four categories. [i] Geometric descriptors reflect dimensions such as bond lengths, bond angles and interatomic distances within the analogue series. [ii] Electronic descriptors represent properties such as atomic charge densities, molecular dipoles, and energy of the highest occupied molecular orbital. [iii] Topological descriptors encode aspects of molecular shape and branching and are frequently represented by graph theory indices. [iv] Physicochemical descriptors include values such as the octanol-water partition coefficient and reflect properties related to the ability of the molecules to traverse biological barriers [11]. Based upon a data array which correlates descriptors with biological activities for a given analogue series of drug-like molecules, the minimal descriptor set capable of differentiating between biological activity and inactivity is determined [12–16]. As a corollary to this, it is possible to deduce the bioactive face of the drug molecule, thereby identifying the pharmacophore.

Of the four classes of descriptor, topological descriptors describe molecular branching and complexity through an exploitation of molecular connectivity [13,17,18]. In chemical graph theory as applied to drug design, a topological index is any of several arithmetically trivial numerical parameters (which are usually graph invariant) of a graph which characterize its topology. Drug molecules may be represented as molecular graphs in which graph vertices correspond to atoms and graph edges represent covalent bonds between atoms; these graphs are typically hydrogen suppressed. The graph model of drug structure describes the bonding pattern of atoms, without explicit reference to overall molecular geometry. Topological graph theory indices are significant because, despite their mathematical simplicity, they are able to successfully differentiate active from inactive biomolecules. Accordingly, such topological indices are immensely useful in efficiently screening large *in silico* drug molecule libraries without the need for detailed molecular orbital calculations.

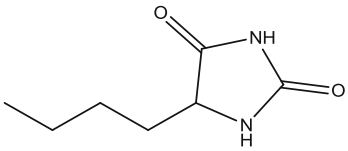
To further broaden the practical utility of these simple indices we have sought to identify hybrid descriptors which are trivial to calculate but which capture data from at least two categories of descriptors. An electrotopological (ET) descriptor combines electronic information and molecular topology to describe the structure at an atomic level. The ET is based solely on atomic connectivity information obtained from the graph without any input from the geometry.

After a series of explorations of various descriptors, we have introduced the following descriptor (ET_z) for the evaluation of neuroactive drug-like molecules:

$$ET_z = \sum_{i=1}^{n} b_i Z_i$$

where Z is the atomic number (number of protons in the nucleus; hydrogen is excluded), and b is the number of covalent bonds from each non-hydrogen atom

in the drug molecule to other non-hydrogen atoms. A representative calculation is as follows:



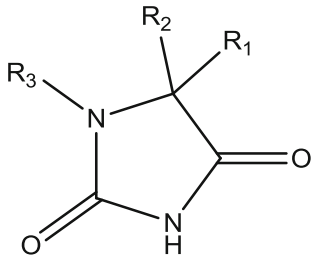
$ET_z = 8(2) + 8(2) + 7(2) + 7(2) + 6(4) + 6(4) + 6(3) + 6(2) + 6(2) + 6(2) + 6(1) = 168$

To demonstrate the utility of this descriptor, we have applied it to a standard series of 15 hydantoin anticonvulsants synthesized and biologically evaluated in our laboratory (See Table 1) [19]. Activities are expressed on a scale of 0–4 where 0 = inactive and 4 is the activity of diphenylhydantoin (phenytoin, a widely used clinically employed therapeutic agent); activities were determined using a rodent pilocarpine-induced seizure model (Fig. 1).

Linear regression on the data of activity levels and ET_z indices gives the following relationship:

$$\text{Anticonvulsant Activity} = 0.0179(ET_z) - 2.201 \; [R^2 = 0.95]$$

Table 1 Activity and ET_z values for 15 hydantoin analogue anticonvulsants



R_1	R_2	R_3	ET_z	Activity
H	H	H	120	0
CH_3	H	H	132	0
CH_2CH_3	H	H	144	0
$CH_2CH_2CH_2CH_3$	H	H	168	1
$C(CH_3)_3$	H	H	168	1
C_6H_5	CH_3	H	252	3
C_6H_5	C_6H_5	H	360	4
C_6H_5	$C(CH_3)_3$	H	288	3
$C(CH_3)_3$	$C(CH_3)_3$	H	216	2
C_6H_5	C_5H_5N	H	363	4
$C(CH_3)_3$	$CH_2CH_2CH_2CH_3$	H	216	2
CH_3	CH_3	CH_3	157	0
C_6H_5	H	H	240	2
H	H	C_6H_5	241	2
$C(CH_3)_3$	CH_3	H	180	1

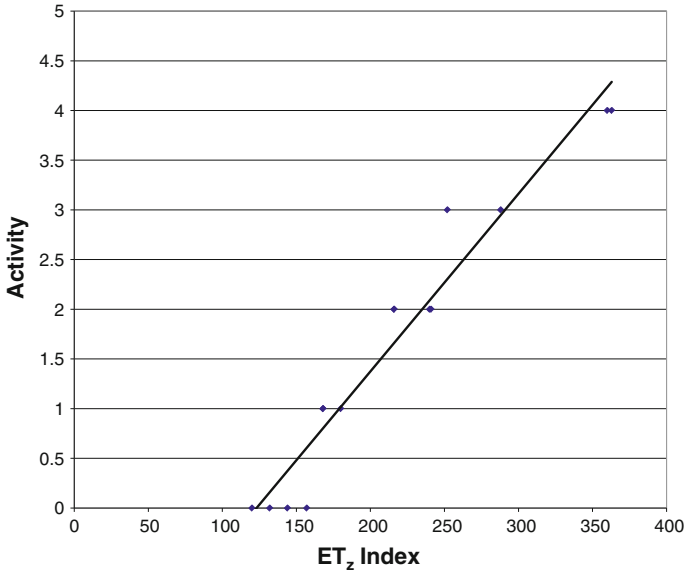


Fig. 1 Activity versus ETz index in 15 hydantoin anticonvulsants

Table 2 Relationship between activity and ET_z index

Activity level	ETz index
0	≤157
1	168–180
2	216–241
3	252–288
4	≥360

Also from our hydantoin test set, the following could be noted (Table 2).

Using the simple ET_z descriptor as the only descriptor, it has been possible to devise a QSAR algorithm with which to predict the anticonvulsant activity of hydantoins (R^2=0.95). In addition, to validating the ET_z descriptor against this standard set of hydantoin anticonvulsants, we also obtained similar statistically significant results for standard sets of barbiturate and acyclic ureide anticonvulsants [19–21]. In our hands, the ET_z descriptor works best for pharmacologically active molecules with molecular weight less than 450 g/mol.

Currently, there are several hundred descriptors available for 2D-QSAR studies [15,16]. A rigorous QSAR study employs multiple descriptors in an unbiased fashion and does not rely on particular descriptors [2]. The ET_z descriptor endeavours to be competitive in this crowded descriptor space by combining both topological and electronic properties (in an extremely computationally efficient manner), rather than being exclusively only topological, electronic or physicochemical. However, it is anticipated that multiple descriptors will usually be required for the development of a statistically significant QSAR prediction algorithm.

Arguably, more physically-based approaches, such as 3D-QSAR, could be used to obtain higher level data for comparison with pharmacological or biological efficacies [22]. 3D-QSAR refers to the application of empirical molecular mechanics calculations using a force field potential function and requiring three-dimensional structures, typically based on molecule superimposition. 3D-QSAR is concerned with overall molecular structure rather than isolated substituent effects; it examines the hydrophobic regions, the steric fields and the electrostatic fields, and the resulting data space is then subjected to dimensionality reduction. Although such physically-based approaches provide additional data, their use is more time demanding. For small data sets (including the 15 hydantoin analogues evaluated in this study), 3D-QSAR offers definite advantages; for larger or more molecularly diverse data sets, 2D-QSAR employing proficient descriptors, such as ET_z, offers almost comparable quality in a more time efficient manner. Moreover, since high-throughput screening of compound libraries (virtual or real) against druggable targets is increasingly being used to discover a wide range of therapies, and brain therapeutics specifically, it is crucial to ascertain if such screening methods, when coupled to QSAR, can adequately explore "neurotherapeutic space" (*i.e.* the total number of molecules that are or could be neuroactive drugs) [23]. Since the size of neurotherapeutic space alone has been conservatively estimated to contain 6×10^{15} molecules [23], the need for simple molecular descriptors, such as ET_z, persists.

In conclusion, an extremely simple descriptor (ET_z) has been introduced which combines data concerning charge number of the nucleus with covalent bonding connectivity. This single simple descriptor is significantly able to differentiate activity from inactivity in a small set of hydantoin anticonvulsants. This emphasizes the power and utility of arithmetically trivial descriptors in QSAR analyses. Although it should not be used in isolation, this hybrid descriptor could be employed in conjunction with a battery of other descriptors as part of a 2D-QSAR approach to drug molecule optimization.

Acknowledgments DFW acknowledges salary support from a Canada Research Chair, Tier 1, in Neuroscience. This work was supported by an operating grant from the Canadian Institutes of Health Research. Neither author declares any conflict of interest.

References

1. P. Gramatica, QSAR Comb. Sci. **26**, 694 (2007)
2. A. Tropsha, P. Gramatica, V.J. Gombar, QSAR Comb. Sci. **22**, 69 (2003)
3. X. Dong, J.O. Ebalunode, S.Y. Yang, W. Zheng, Curr. Comput. Aided Drug Des. **7**, 18 (2011)
4. R.C. Braga, C.H. Andrade, Mini Rev. Med. Chem. **12**, 573 (2012)
5. N.L. Kruhlak, R.D. Benz, H. Zhou, T.J. Colatsky, Clin. Pharmacol Ther. **91**, 529 (2012)
6. J. Sutherland, L.A. O'Brien, D.F. Weaver, J. Med. Chem. **47**, 5541 (2004)
7. J. Sutherland, D.F. Weaver, J. Comput. Aided Mol. Des. **18**, 309 (2004)
8. M.Y. Connolly, WIREs Comput. Mol. Sci. **2**, 435 (2012)
9. P. Riera-Fernández, R. Martín-Romalde, F.J. Prado-Prado, M. Escobar, C.R. Munteanu, R. Concu, A. Duardo-Sanchez, H. González-Díaz, Curr. Top. Med. Chem. **12**, 927 (2012)
10. C.R. Munteanu, E. Fernández-Blanco, J.A. Seoane, P. Izquierdo-Novo, J.A. Rodríguez-Fernández, J.M. Prieto-González, J.R. Rabuñal, A. Pazos, Curr. Pharm. Des. **16**, 2640 (2010)
11. R. Gozalbes, A. Pineda-Lucena, Comb. Chem. High Throughput Screen. **14**, 548 (2011)
12. D.T. Stanton, Curr. Comput. Aided Drug Des. **8**, 107 (2012)

13. J. Gálvez, M. Gálvez-Llompart, R. García-Domenech, Expert Opin. Drug Discov. **7**, 133 (2012)
14. R.D. Combes, Adv. Exp. Med. Biol. **745**, 96 (2012)
15. R. Gozalbes, A. Pineda-Lucena, Comb. Chem. High Throughput Screen. **14**, 548 (2011)
16. H. Timmerman, R. Todeschini, V. Consonni, R. Mannhold, H. Kubinyi, *Handbook of Molecular Descriptors* (Wiley-VCH, Weinheim, 2002)
17. J. Gálvez, R. García-Doménech, Curr. Comput. Aided Drug Des. **6**, 252 (2010)
18. K. Roy, I. Mitra, Comb. Chem. High Throughput Screen. **14**, 450 (2011)
19. J. Sutherland, D.F. Weaver, J. Chem. Inf. Comput. Sci. **43**, 1028 (2003)
20. J. Bikker, J. Kubanek, D.F. Weaver, Epilepsia **35**, 411 (1994)
21. M. Khalil, D. Weaver, J. Pharm. Pharmacol. **42**, 349 (1990)
22. L. Zhang, K.C. Tsai, L. Du, H. Fang, M. Li, W. Xu, Curr. Med. Chem. **18**, 923 (2011)
23. D.F. Weaver, C.A. Weaver, J. Pharm. Pharmacol. **63**, 136 (2011)