ON CHARACTERIZATION OF PHARMACOPHORE

Milan Randić
Department of Mathematics and Computer Science
Drake University, Des Moines, IA 50311
e-mail: milan.randic@drake.edu

Received 6.12.1999

Abstract

We consider use of local graph invariants for characterization of molecular fragments that are responsible for the dominant features of structure-activity relationship.

Introduction

Mathematical descriptors of molecular structure, such as various topological indices[1], have been widely used in structure-property-activity studies, including the multiple regression analysis (MRA), the principal component analysis (PCA), the pattern recognition, the artificial neural networks (ANN), optimization of lead compound, search of combinatorial libraries and combinatorial optimization of the lead compound, and the similarity-dissimilarity studies.

In this article we will focus attention on use of molecular descriptors for characterization of pharmacophore. The following steps are involved in such applications:

(1) Selection of local descriptors;

(2) Selection of potential fragment;

(3) Similarity/dissimilarity study;

(4) Construction of partial order;

(5) Selection of new fragment…

i. e., repeat of the steps (2) - (4) till a satisfactory partial order is obtained.

In a search for unknown pharmacophore one may start by selecting a molecular fragment such as a smallest common part to all active compounds, or a disjointed group of atoms common to the most active compounds investigated, and examine if a similarity among molecules based on so selected fragment correlates with the relative activities of the molecules considered. Once a promising fragment has been identified one can continue and augment such fragment by including additional neighboring atoms. Two or

three most active compounds should be selected as the standards and similarity of all other compounds with respect to the standards should be examined. Such data will allow one to extract a partial order for the compounds considered which represents common ranking of all the compounds relative to the selected standards. If such ordering parallels the relative activities the common fragment can be viewed as the sought pharmacophore. If not, the process is repeated by considering other molecular fragments till a satisfactory parallelism is established between the relative similarities and the relative activities.

In this paper we will illustrate the approach by focusing attention on a 7-atom pharmacophore in dozen compounds closely related to methyl-2-oxypropylnitrosamine (MOP) which shows an unusually high mutagenicity. The compounds that apparently are all similar to each other differ in their mutagenic activity almost by three orders of magnitude. By using atomic ID numbers[2] as molecular descriptors Randic and collaborators[3] identified a seven atom fragment that offered a parallelism between the local similarities of nitrosamines relative to MOP and the relative mutagenicities as the mutagenic pharmacophore for this class of nitrosamines. We will adopt the already recognized seven-atom molecular fragment as the pharmacophore and will focus attention to a search for alternative molecular descriptors that may produce similar results. There are essentially two reasons for considering alternative molecular descriptors: (1) Novel descriptors may lead to better regression analysis, as has recently been illustrated with use of molecular descriptors which involve variable parameters to be optimized during the regression procedure,[4-8] and (2) Alternative descriptors may be computationally simpler which is an important factor when screening combinatorial libraries which may contain several hundred thousands of structures for which molecular descriptors have to be evaluated.

There are hundreds of available mathematical descriptors for molecular graphs, many of which can be computed by programs such as MOLCONN, POLLY, CODESSA, TSAR, TAM. Even though in this study we are not considering 3D molecular structures, let us mention that several of these indices have been extended to characterization of 3-dimensional molecular structure.[9] For our present study it is more important that many of the indices allow one to construct local molecular descriptors. When one is interested in

comparison of local molecular features local molecular indices ought to be employed. In the following section we will review selected local molecular descriptor.

Local molecular descriptors

Of the three "classical" topological indices, the Wiener number,[10] the Hosoya Z topological index[11] and the connectivity index[12] only the last, being bond additive, immediately offers characterization of local atomic environment. Such characterization can be obtained simply by summation of the contributions that are limited to the bonds of the selected molecular fragment. Recently partitioning of the Wiener number[13,14] and the Hosoya Z topological index[14] was outlined. Such partitions of global molecular descriptors into bond contribution permits construction of the local description useful for the search of pharmacophores.

Several atomic descriptors have been reported in the literature. One of the first atomic descriptor was suggested by Kier by applying the algorithm used for construction of the connectivity index[12] and the higher order connectivity indices[15] to a single atom, leading to the so called "zero order" connectivity index.[16] This index has been found useful in combination with other connectivity indices in many multiple regressions but it has been rarely used alone, because of its high degeneracy. The degeneracy is in this case a consequence of the dependence of the zero order connectivity index solely on the distribution of valence among atoms present. Atomic ID numbers[2] were suggested as an alternative because they displayed very low degeneracy and hence show a remarkable high degree of discriminatory power among atomic environments. They are based on weighted paths, where the weights of individual paths are given in an analogous way to determination of the path contributions to the higher order connectivity indices. The atomic ID (identification number) is given as the sum of the contributions of all weighted paths in a molecule that originate at the atom considered. Typically terminal atoms have smaller number while more centrally located atoms and atoms having greater valence will also have greater atomic ID number. Additional atomic descriptors, or local vertex invariants (LOVI) as they are referred to by Balaban,[17] have been reported in last couple

of years. For every matrix that can be associated with a graph one can consider its row sums as novel local atomic descriptors.

A number of novel graph matrices have been introduced or resurrected in recent years that include for instance the Wiener matrix,[18] the Hosoya Z matrix,[19] the Restricted Random Walk matrix,[20] the Distance/Distance matrix,[21] the Resistance-Distance matrix,[22] the Detour matrix,[23] two kinds of Path matrices,[24, 25] etc. In this way recently a dozen novel atomic descriptors were generated.

We will examine one particular local invariant, the augmented valence, that has only recently been introduced for characterization of molecular complexity.[26] As we will see this particular descriptor, considered in the next section, has some apparent advantages: it can be easily computed, and it can be easily modified. Both these features are important when one think of use of such descriptors for searching combinatorial libraries.

Augmented valence

A simple characterization of an atomic environment is a list of numbers of nearest neighbors, the next nearest neighbors etc., n_l, n_2, n_3, . . . In the case of trees this list is identical to the list of paths of different length: p_1, p_2, p_3, . . . Such list can be converted to a single entry by summing the members of the sequence with appropriate weights: $w_1 n_l + w_2 n_4 + w_3 n_3 +$. . . Another simple list: s_l, s_2, s_3, . . . is given by the sum of valences of neighbors at increasing distance. Again a single descriptor can be obtained from the sequence by constructing a weighted sum: $w_1 s_1 + w_2 s_2 + w_3 s_3 +$. . It is plausible to assume that more distant neighbors will have lesser effect on the atom under consideration. Hence, the simple weighting algorithm $w_k = (1/2)^k$ offers novel local atomic invariant. In Table 1 we illustrate so constructed new atomic invariants for carbon atoms of 1-methylpentane (assuming the standard numbering for carbon atoms):

Table 1. Contributions to the augmented valence for 6 carbon atoms in 1-methylpentane; weights calculated as $w_k=(1/2)^k$.

atom	contributions
1	1 +3/2+3/4+2/8+1/16=3.5625
2	3+4/2+2/4+ 1/8=5.6250
3	2+5/2+3/4=5.2500
4	2+3/2+3/4+2/8=4.5000
5	1 +2/2+2/4+3/8+2/16=3.0000
6	1 + 3/2 + 3/4 +2/8 +1/ 16 = 3.5625

The successive numerators in so constructed "augmented" valence and the sums of the valences of carbon atoms at increasing distance from the atom considered white the denominators are the successive powers of two. If we add all atomic contributions we obtain augmented valence for the molecule, $\xi\xi$, which for 1-methylpentane gives 25.5000. If we add only the contributions of symmetry non equivalent atoms, which in this case gives 21.9375, we obtain the molecular complexity index ξ.[26]

Although the $(1/2)^k$ distance dependence may be viewed as a "short range" when compared to 1/ k distance dependence, nevertheless even the 10-th shell of neighbors still will influence the magnitude of the augmented valence at the fourth decimal place (1/1024 being approximately 0.0001). We decided therefore to further curtail the role of more distant neighbors by using the reciprocal factorials as the weights, thus assuming $w_k = 1/k!$. For this new weighting modification we obtain the revised atomic augmented valence illustrated in Table 2 again on 1-methylpentane:

Table 2. Contributions to the augmented valence for 6 carbon atoms in 1-methylpentane; weights calculated as $w_k=1/k!$.

atom	contributions
1	1+ 3/2+ 3/6+2/24+ 1/120 = 3.0917
2	3+4/2+2/6+ 1/24 = 5.3750
3	2+ 5/2+ 3/6 = 5.0000
4	2+3/2+3/6+ 2/24 = 4.0833
5	1+ 2/2+ 2/6+ 3/24+ 2/120 = 2.4750
6	1+3/2+3/6+2/24+1/120 = 3.0917

Again when we add all atomic contributions we obtain novel molecular descriptor (23.1167). Although we have almost halved the range of the neighbors that could make a significant contribution to atomic environment the relative magnitudes of the revised and the previous atomic descriptors have little changed. Again terminal atoms have the smallest values for the augmented valence, while central atoms, in particular those associated with higher valence, have the largest values for the augmented valence. In the following section we will illustrate use of the augmented valence descriptors for local atomic environment.

Characterization of the pharmacophore, the critical structure

We consider dozen mutagenic compounds closely related to methyl-2-oxypropylnitrosamine (MOP). Their relative mutagenicity has been reported in the literature[27] are listed in Table 3. Randic et al.[3] searched for molecular fragment that could be responsible for the mutagenicity of these compounds. After examining several molecular fragments they concluded that a seven-atom fragment common to the dozen compounds act as a pharmacophore. This finding was based on the fact that for this particular fragment (and not the alternatives considered) the relative mutagenicities parallel the relative similarity between the seven-atom fragment in MOP and the seven-atom fragments in other structures. Atomic descriptors used were the atomic ID numbers.

In Table 3 we give the abbreviated names of the compounds (as reported in the study of Langebach et al.,[27] their relative mutagenicity followed by the relative similarity towards MOP when descriptors used were the atomic ID numbers, the augmented valence based on factorial and decimal weights.

Table 3. Relative dissimilarity of 12 mutagenic compounds compared to MOP calculated for three different descriptors.

	mutagenicity	atom ID	1/n!	$1/10^n$
MOP	650	0	0	0
MHP	380	0.209	1.785	1.111
BOP	250	0.428	3.767	1.437
2 MOB	210	0.377	1.655	1.106
MP	105	0.548	3.571	2.222
HPOP	90	0.482	3.066	1.327
POP	75	0.502	2.353	1.216
3-MOB	30	0.80'7	3.407	3.608
2-HPP	25	0.630	2.748	1.644
DP	20	0.616	4.007	2.529
BHP	10	0.546	3.342	1.727
3-HPP	1	0.632	3.248	2.578

Use of weights based on powers of 10 allow one to convert the list of the nearest neighbors for each atom immediately as the augmented valence. For example, for atom 1 in MOP the list of the nearest neighbors is: 2, 3, 3, 3, 4, 3, i. e., atom 1 has two nearest neighbors, three next nearest neighbors, etc. to obtain the augmented valence for atom 1 based on decimal weights just convert the above sequence into a single decimal number: 2.33343. This was the number used in the evaluation of similarities/dissimilarities shown in the last column of the above table.

In the table below we summarized the results of a quadratic regression using the three sets of descriptors. Here *r*, *s*, and *F* represent the coefficient of the regression, the standard error, and Fisher ratio, respectively. Clearly atomic ID numbers as local descriptors are visibly better than the descriptors derived by the augmenting valences. However, computationally the atomic ID are more difficult to obtain because they are based on the enumeration of paths (which itself is a problem of NP computational complexity, the concept introduced by Karp [28]).

Table 4. Results of a quadratic regression model for mutagenicity of MOP related compounds using three different sets of descriptors.

	atom ID	1/n!	$1/10^n$
r	0.9769	0.8931	0.9033
s	46.0	96.8	92.3
F	94	18	20

Concluding Remarks

Search for better molecular descriptors for use in MRA is of considerable interest for the following reasons:

(1) Interpretation: Since MRA does not imply cause-effect relationship use of alternative descriptors may point to other structural factors that parallel a particular molecular property. This may help interpretation of the results.

(2) Accuracy: Novel descriptors can dramatically reduce the standard error of a regression which could possibly allow detection of an experimental error that is currently hidden in a scatter of experimental and calculated points.

(3) Speed: The speed of calculation need not be important when considering smaller molecules or smaller number of molecules. However, this is no longer the case when one screens combinatorial libraries that may involve 100,000 compounds or more. In such applications computational complexity of molecular descriptors becomes a factor.

It is this last point that motivated us to seek alternative descriptors to atomic ID numbers. Although the augmented valence shows some limitations we hope that similarly modified descriptors may show better parallelism with atomic ID numbers. A quadratic regression between the augmented valence (using decimal weights) and atomic ID has the coefficient of regression r = 0.9257, the standard error s = 0.39, and the Fisher ratio F = 27. This correlation is encouraging, even though the resulting augmented valences were not as successful in the characterization of the 7-atom pharmacophore.

References and Notes

1. M. Randić, Topological Indices. *Encyclopedia of Computational Chemistry* (P. von Rague Schleyer, Editor-in-Chief, London: Wiley, 1998, pp.3018-3032.
2. M. Randić, *Molecular Basis of Cancer, Part A: Macromolecular Structure, Carcinogens, and Oncogens,* (R. Rein, Ed.), Alan Liss Publ., 1985, pp.309-318.
3. M. Randić, B. Jerman-Blazic, D. H. Rouvray, P. G. Seybold, S. C. Grossman, *Int. J. Quantum Chem: Quantum Biol. Symp.* **1987**, *14*, 245.
4. M. Randić, *Chemom. Intell. Lab. Syst.* **1991**, *10*, 213.
5. M. Randić, *J. Comput. Chem.* **1991**, *12*, 970.
6. M. Randić, J. Cz, Dobrowolski, *Int. J. Quantum Chem: Quantum Biol. Symp.* **1998**, *70*, 1209.
7. M. Randić, S. C. Basak, J. *Chem. Inf. Comput. Sci.* **1999**, *39*, 261.
8. M. Randić, M. Pompe, SAR & *QSAR* **1999**, *10*, 451-457.
9. M. Randić, M. Razinger, On Characterization of 3D Molecular Structure. *From Topology to Three-Dimensional Geometry* (A. T. Balaban, Ed.), New York: Plenum Press, 1997.
10. H. Wiener, J. *Am. Chem. Soc.***1947**, *69*, 17.
11. H. Hosoya, *Bull. Chem. Soc.* Jpn.**1971**, *44*, 2332.
12. M. Randić, J *Am. Chem. Soc.* **1975**, *97*, 6609.
13. I. Lukovits, *Quant. Struct.-Act. Relat.* **1990**, *9*, 227.
14. M. Randić, J. Zupan, *New J. Chem.* (submitted).
15. L. H. Kier, W. J. Murray, M. Randić, L. H. Hall, *J Pharm. Sci.* **1976**, *65*, 1226.
16. L. B. Kier, L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, New York: Academic Press. 1976.
17. A. T. Balaban, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 398.
18. M. Randić, *Chem. Phys. Lett.* **1993**, *221*, 478.
19. M. Randić, *Croat. Chem. Acta* **1994**, *67*, 415.
20. M. Randić, *Theor. Chim. Acta* **1995**, *92*, 97.
21. M. Randić, A. F. Kleiner, L. M. DeAlba, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 277.
22. D. J. Klein, M. Randić, *J. Math. Chem.* **1993**, *12*, 81.
23. D. Amic, N. Trinajstić, *Croat. Chem. Acta* **1995**, *68*, 53.
24. M. Randić, D. Plavšič, M. Razinger, *MATCH* **1997**, *35*, 243.
25. O. Ivanciuc, A. T. Balaban, *MATCH* **1994**, *30*, 141.
26. M. Randić, *Croat. Chem. Acta* (submitted).
27. R. Langebach, C. Kruszynski, R. Gingell, T. Lawson, D. Nagel, P. Pour, S. C. Nesnow, Chapter 16 in *Structure-Activity Correlation as a Predictive Tool in Toxicology* (L. Goldberg, Ed.), New York: McGraw-Hill, 1983.
28. R. M. Karp, Reducibility among combinatorial problems. In: *Complexity of Computer Computations* (R. E. Miller and J. W. Tatcher, Eds.), Plenum Press, 1972, pp. 85-103.

Povzetek

Predlagamo uporabo lokalnih invariant grafov za označitev molekulskih fragmentov, ki so odgovorni za dominantne lastnosti v odnosu struktura-aktivnost.