On Characterization of Pharmacophore

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

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

Introduction

Mathematical descriptors of molecular structure, such as various topological indices (Randic 1998), 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 the steps (2) - (4) till satisfactory partial order is obtained.

In practice one select as the first fragment a smaller common part to all active compounds, and later augment it by including neighboring atoms. Two or three most active compounds are selected as standards and the similarity of all other compounds with respect to the standards is found. Such data allow one to extract the partial order which represents common ranking of all compounds relative to the selected standards. If such ordering parallels relative activities the common fragment can be viewed as pharmacophore. If not, the process is repeated by considering other molecular fragments till satisfactory parallelism is established between the relative similarities and relative activities.

In this paper we will focus attention on characterization of a 7-atom pharmacophore in dozen compounds closely related to methyl-2-oxypropylnitrosamine (MOP) which has displayed unusually high mutagenicity. By using atomic ID numbers Randic and collaborators (Randic et al. 1987) found a seven atom fragment that offered a parallelism between the local similarities of nitrosamines relative to MOP and the relative mutagenicities that covered three orders of magnitude in relative values.

We will adopt the already recognized seven-atom molecular fragment as the pharmacophore and will focus attention of search for suitable molecular descriptors that may produce similar results. There are hundreds and hundreds of reported mathematical descriptors for molecular graphs, many of which can be computed by programs such as MOLCONN, POLLY, CODESSA, TSAR, TAM. Several of these indices have been extended to characterization of 3-dimensional molecular structure (Randic and Razinger, 1997). Several 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 (Wiener 1947), the Hosoya Z topological index (Hosoya 1971) and the connectivity index χ (Randic 1975) only the last, being bond additive, immediately offers characterization of local atomic environment when summation of the contributions is limited to bonds of selected molecular fragment. Recently partitioning of the Wiener number (Lukovits 1990, Randic and Zupan 1999) and the Hosoya Z topological index (Randic and Zupan 1999) was outlined. Such partition of global molecular descriptor into bond contribution permits construction of the corresponding local descriptors.

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 and the higher order connectivity indices to a single atom, leading to the so called zero order connectivity index (Kier and Hall 1976). 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 the consequence of the dependence of the zero order connectivity index solely on the distribution of valence among atoms present. Atomic ID numbers were suggested (Randic 1985) as an alternative because they displayed remarkable degree of discriminatory power. 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 simply the sum of the contributions of all weighted paths in a molecule that originate at the atom considered. Typically terminal atoms have smaller ID number while more centrally located atoms and atoms having greater valence will also have greater atomic ID number. Many additional atomic descriptors, or local vertex invariants (LOVI) as they are referred to by Balaban, have been reported in last couple of years (Balaban 1994). Each time a matrix can be associated with a graph the row sums give the corresponding LOVI. A number of novel graph matrices have been introduced or resurrected in recent years that include for instance the Wiener matrix (Randic 1993), the Hosoya Z matrix (Randic 1994), the Restricted random walk matrix (Randic 1995), the Distance/Distance matrix (Randic, Kleiner and DeAlba 1994), the Resistance-distance (Klein and Randic 1993), the Detour matrix (Ivanciuc and Balaban 1994, Amic and Trinajstic 1995), Path matrices (Randic, Plavsic and Razinger, 1997). 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 (Randic 1999). 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 feature are important when one think of possible application of such descriptors in combinatorial libraries.

Augmented valence

A simple characterization of an atomic environment would be listing of number of nearest neighbors, the next nearest neighbors etc., n1, n2, n3, . . . Such list can be converted to a single entry by summing the members of the sequence with appropriate weights: w1 n1 +w2 n2 + w3 n3 + . . . It is plausible to assume that more distant neighbors will have lesser effect on the atom under consideration. Hence, the simple weighting algorithm wk = 1/2^k offers novel local atomic invariant (Randic 1999). In the Table 1 we illustrate the new atomic invariants for carbon atoms of 1-methylpentane:

Table 1

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

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 the contributions of symmetry non equivalent atoms, which in this case gives 21.9375, we obtain the molecular complexity index ξ (Randic 1999).

Although the 1/2^k distance dependence may be viewed as "short range" if compared to 1/k distance dependence, nevertheless the 10-th shell of neighbors still will influence 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 wk = 1/k! With this new weighting modification we obtain the revised atomic augmented valence illustrated in Table 2 again on 1-methylpentane:

Table 2

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 substructure

We consider dozen mutagenic compounds closely related to methyl-2-oxypropylnitrosamine (MOP). Their relative mutagenicity as been reported in the literature (Langebach et al 1983) shows span of two orders of magnitude, the most potent being MOP. Randic et al (1987) examined alternative molecular fragments as the source for shown mutagenicity and concluded that a seven-atom fragment common to the dozen compounds examined can be assigned as pharmacophore. This finding was based on the fact that for this fragments (and not few alternatives also considered) the relative mutagenicities parallel the relative similarity between the 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), their relative mutagenicity followed by the relative similarity towards MOP when using as descriptors the atomic ID numbers (as reported by Randic et al., 1987), the augmented valence based on decimal weights (to be outlined later).

Table 3

	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	0.807	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

In the following Table we summarized the results of simple regression using the three sets of descriptors. We are interested in comparison between different descriptors and the results are based on a quadratic fit.

Table 4

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

Clearly atomic ID numbers as local descriptors are visibly better than the alternative descriptors derived by augmenting valences. Computationally atomic ID are more difficult to obtain because they are based on the enumeration of paths (which itself is a problem of NP complexity).

Use of 1/n! weights is not an optimal solution when constructing local atomic descriptors based on augmenting valence. We examined as an alternative weighs given by 1/10^n, which attenuates the influence of the nearest neighbors, the next nearest neighbors, etc. for first dozen shells more dramatically. As we see from Table 4 this improves the regression statistics.

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. Incidentally, use of weights based on powers of 10 allow one to convert the list of the nearest neighbors for each atom immediately as the value of 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 number: 2.33343. This was the number used in the evaluation of similarities /dissimilarities shown in the last column of Table 3.

References

Amic, D.; and Trinajstic, N. 1995. On the Detour Matrix. *Croat. Chem. Acta,* 68: 53-62.

Balaban, A. T. 1994. Local versus Global (i. e. Atomic versus Molecular) Numerical Modeling of Molecular Graphs, *J. Chem. Inf. Comput. Sci.,* 34: 398-402.

Hosoya, H. 1971. Topological Index. A Newly Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons. *Bull. Chem. Soc. Jpn.,* 44: 2332-2339.

Ivanciuc, O.; and Balaban, A. T. 1994. Design of Topological Indices. part 8. Path Matrices and Derived Molecular Invariants, *MATCH,* 30: 141-152

Kier, L. B.; and Hall, L. H. 1976. *Molecular Connectivity in Chemistry and Drug Research,* New York: Academic Press.

Klein, d. J.; and Randic, M. 1993. Resistance Distance, *J. Math. Chem.* 12: 81-95.

Langebach, R.; Kruszynski, C.; Gingell, R.; Lawson, T.; Nagel, D.; Pour, P.; and Nesnow, S. C. 1983. Chapter 16 in *Structure-Activity Correlation as a Predictive Tool in Toxicology* (L. Goldberg, Ed.), New York: McGraw-Hill.

Lukovits, I. 1990. Wiener indices and Partition Coefficients of Unsaturated Hydrocarbons, *Quant. Struct.-Act. Relat.,* 9: 227-231.

Randic, M. 1975. On Characterization of Molecular Branching, *J. Am. Chem. Soc.,* 97: 6609-6615.

Randic, M. 1985. Graph Theoretical Approach to Structure-Activity Studies: Search for Optimal Antitumor Compounds. *Molecular Basis of Cancer, Part A: Macromolecular Structure, Carcinogens, and Oncogens,* (R. Rein, Ed.), Alan Liss Publ. pp. 309-318.

Randic, M.; Jerman-Blazic, B.; Rouvray, D. H.; Seybold, P. G.; and Grossman, S. C, 1987. The Search for Active Substructures in Structure-Activity Studies, *Int. J. Quant. Chem.: Quant. Biol. Symp.,* 14: 245-260.

Randic, M. 1993. Novel Molecular Descriptor for Structure-Property Studies, *Chem. Phys. Lett.,* 211: 478-483.

Randic, M. 1994. Hosoya Matrix - A Source of New Molecular Descriptors, Croat. Chem. Acta, 67: 415-429.

Randic, M.; Kleiner, A. F.; DeAlba, L. M. 1994. Distance/Distance matrices, *J. Chem. Inf. Comput. Sci.,* 34: 277-286.

Randic, M. 1995. Restricted Random Walks, *Theor. Chim. Acta,* 92: 97-106.

Randic, M.; and Razinger, M. 1997. On Characterization of 3D Molecular Structure. *From Topology to Three-Dimensional Geometry* (A. T. Balaban, Ed.), New York: Plenum Press.

Randic, M.; Plavsic, D.; and Razinger, M. 1997. Double Invariants, *MATCH,* 35: 243-259.

Randic, M. 1998. Topological Indices. *Encyclopedia of Computational Chemistry* (P. von Rague Schleyer, Editor-in-Chief), London: Wiley.

Randic, M. 1999. On the Concept of Molecular Complexity, *Croat. Chem. Acta* (submitted).

Randic, M.; and Zupan, J. 1999. On the Interpretation of the Well Known Topological Indices, *New J. Chem.* (submitted).

Wiener, H. 1947. Structural Determination of Paraffin Boiling Points, *J. Am. Chem. Soc.,* 69: 17-20.