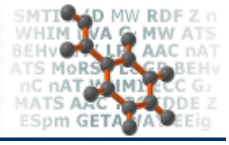
Molecular Descriptors
the free online resource



Basic requirements for valid molecular descriptors

Viviana Consonni and Roberto Todeschini
Milano Chemometrics and QSAR Research Group - Dept. of Environmental Sciences, University of Milano-Bicocca, P.za della Scienza 1 – 20126 Milano (Italy)

Several scientists are involved in searching for new molecular descriptors able to catch new aspects of the molecular structure. This kind of research involves creativity and imagination together with solid theoretical basis to obtain numbers with some structural chemical meaning. "There are no restrictions on the design of structural invariants, the limiting factor is one's own imagination." [1].
A molecular descriptor can be more or less useful, simple, interpretable, etc., but, in any case, it has to fulfil some mathematical requirements.
In particular, the basic properties a molecular descriptor MUST HAVE are:

1. invariance with respect to labelling and numbering of the molecule atoms
2. invariance with respect to the molecule roto-translation
3. an unambiguous algorithmically computable definition
4. values in a suitable numerical range for the set of molecules where it is applicable to

Point 1 requires that the value of a molecular descriptor does not depend (is invariant) on how the molecule atoms are lebelled or numbered. Descriptors which make use of the atom numbering in their definition have to use some canonical unique numbering based on unequivocal rules.
Point 2 requires that the value of a molecular descriptor does not depend on the absolute values of numerical coordinates defining the atom positions with respect to some arbitrary origin. For example, a descriptor value cannot have different values depending on the position of the molecule with respect some fixed reference axis.
According to point 3, a molecular descriptor must be defined by a computable mathematical expression whose terms have to be not ambiguous and clearly available from the molecular structure.
According to point 4, the values of a molecular descriptors must be in an acceptable numerical range, avoiding singular points and values as 10^{12} or 10^{-8}. For example, descriptors defined on the product of some atomic property quickly reach large numerical values for big molecules.
The mathematical rules given above MUST hold for all the molecular descriptors.

Moreover, good molecular descriptors SHOULD HAVE other important characteristics:

a. a structural interpretation
b. a good correlation with at least one property
c. no trivial correlation with other molecular descriptors
d. gradual change in its values with gradual changes in the molecular structure
e. not including in the definition experimental properties
f. not restricted to a too small class of molecules
g. preferably, some discrimination power among isomers
h. preferably, not trivially including in the definition other molecular descriptors
i. preferably, allowing reversible decoding (back from the descriptor value to the structure)

Suggestions about the required characteristics of molecular descriptors are also discussed in M. Randic (1996), *Molecular bonding profiles*, J. Math. Chem., 19, 375-392.