# Molecular Descriptors

* Abstract: - to evaluate chemical properties avoiding chemical synthesis and reducing expensive and time-demanding laboratory testing, it is necessary to build in silico models establishing a mathematical relationship between the structures of molecules and the considered properties.
* Molecular descriptors have two main categories, experimental and theoretical descriptors.
* Introduction: - the overall number of chemicals synthesized, isolated, or marketed is increasing exponentially. The CAS REGISTRY database currently contains more than 93 million unique organic and inorganic chemical substances; it is updated daily and nearly 15,000 new substances are added to it each day.
* Also, in the same time techniques like molecule structure generation and computer assisted structure elucidation (CASE) of chemical compounds give the possibility to design virtual molecules using in silico methods and potentially allow researchers to design millions of molecules in seconds.
* Quantitative structure-activity relationships (QSAR) and the similar quantitative structure-property relationships (QSPRs) are chemoinformatic techniques that allow establishing a relationship between the molecular structure and a defined endpoint.
* It is necessary to define, from a chemoinformatic point of view, what is molecular structure is? The molecular structure is a set of parameters able to describe a molecule in quantitative terms.
* Depending on the selected molecular representation, we can describe a molecule by means of experimental descriptors or theoretical descriptors.
* Experimental descriptors are all the experimental measurements such as octanol-water partition coefficient, molar refractivity, polarizability, and in general, any physicochemical property obtained applying a well-specified experimental procedure.
* The major drives that foster the development of QSAR are essentially two: - first one is the rapid growth of number of available chemicals together with freely accessible databases collecting experimental data.
* The second stimulus comes from institutions. REACH regulation on the registration, evaluation, authorization, and restriction of chemicals in Europe promotes the use of QSAR methodologies to predict not available experimental data.
* Nowadays the analysis is more focused on the recognition of the relationships among theoretical descriptors and the considered experimental property.
* A molecular descriptor must:
  + Be invariant to atom labeling and numbering
  + Be invariant to the molecule roto-translation
  + Be defined by an unambiguous algorithm
  + Have a well-defined applicability on molecular structures
* Also, a molecular descriptor should meet the following requirements:
  + Should have structural interpretation
  + Should have a good correlation with at least one experimental property
  + Should not have trivial relation with other molecular descriptors
  + Should not be based on experimental properties
  + Should preferably be continuous
  + Should preferably show minimal degeneracy
  + Should preferably be simple
  + Should preferably be applicable to a broad class of molecules
  + Should preferably be able to discriminate among isomers
  + Should preferably have calculated values in a suitable numerical range for the set of molecules where it is applicable to
* The first rule can be used to understand if a molecular descriptor is well defined, but they do not permit to realize if a descriptor is useful or not to predict property.
* A molecular descriptor should have an applicability domain; it means that it should be clear to which class of molecules can be applied.
* The recent challenge for the molecular descriptors is that it should face with the possibilities to be applied on different classes of compounds like salts, ionic liquids, peptides, polymers, and nanostructures.
* A molecular descriptor molecular descriptor be a continuous variable and show a low degeneracy; it means that a molecular descriptors should be able to account for variations in molecular structure even if validation is small.
* How to Represent Molecules: - the simplest molecule representation is the chemical formula that codifies the atom types and their occurrences within a molecule.
* The topology representation is the most popular one. It is a two-dimensional representation that, in addition to the atomic composition, includes the information about connectivity of atoms in the molecule in terms of a molecular graph.
* In 3D representation every atom of the molecule is associated with a set of three-dimensional coordinates determining its own position in the three-dimensional space.
* 4D representation takes also into account molecular properties deriving from electron distribution and the interactions of the molecules with probes.
* How to Codify Chemical Structures: - the two main types of chemical files can be distinguished, one concerning the topological representation and the other the geometrical representation of molecules.
* The linear notations mostly use the SMILES file format. The latter includes all chemical file formats that can store the three-dimensional coordinates of provide a standard format to store chemical data based on an XML dialect is the Chemical Markup language (CML).
* Molecular Structure Curation: - the fundamental assumption at this basis of molecular descriptor calculation is the correctness of the molecular structures on which calculations algorithms are applied.
* Data curation is usually composed of different steps to be applied to a whole dataset. Chemical structure curation has been recently detailed and has been split in five main steps:
  + Removal of chemical structures that cannot be treated by the considered application or software, typically inorganics, mixtures, and eventually organometallics
  + Structural conversion and cleaning
  + Standardization and normalization of specific chemotypes
  + Removal of duplicates
  + Final manual checking
* To calculate molecular descriptors chemical structures must be encoded into a specific computational format, such as SMILES, SYBYL, or MDL.
* The third step of the curation procedure tasks into account the fact that a molecule, or even a single fragment, can be represented either in Kekulé or aromatic form; since the representation is different, in principle it can affect the molecular descriptor values even if the chemical information is identical.
* The last step concerns the final check of all the codified molecular structures, which has to be done manually and, consequently, could be tedious and time demanding especially for big datasets.
* Constitutional Descriptors: - Constitutional descriptors are the simplest molecular descriptors that can be calculated starting from a molecular structure. It comprises all those descriptors representing a molecular structure, which take into account only chemical composition and do not encode information about the overall topology and geometry.
* The cyclomatic number (*c*), also known as circuit rank, is the most known ring descriptor and it is defined as:

*c* = |E| - |V| + *D*

* Where |E| is the number of bonds, |V| is the number of atoms, and *D* is the number of disconnected fragments in the molecular structure.
* The cyclomatic number is the minimum number of edges to be removed from a molecular graph in order to remove all its cycles, making it into a forest, that is, an acyclic graph.
* Topological Indices (TIs): - topological indices are two-dimensional descriptors derived from the topological representation of molecular structures; no information spatial distribution of the atoms is considered in the calculation of these descriptors.
* Molecular Graph: - a normal 2D representation of the molecules is molecular graph in which G = (V, E), is a representation of a set V of objects, called vertices, and a set E of links between vertices, called edges.
* The molecular graph is typically an undirected, weighted, sparse graph.
* Application of graph theory on molecular structure permits to handle chemical graphs adopting well-defined algorithms useful to explore structural properties of the molecules.
* A graph theory can be used to extract information from the molecular graph, such as the calculation of matrices depending on graph connections and the definition of graph invariants that ate those properties that depend only on the abstract structure of the graphs and do not depend from graph representations.
* A graph invariant is complete if it has equal values only for topological indices.
* A graph is commonly represented in two standard ways, as a collection of adjacency lists or as an adjacency matrix.
* The adjacency lists are a collection of |V| lists, one for each atom *i*, where every list *Adj[i]* includes the atoms connects to the *i*-th atom together with the corresponding bond orders.
* The adjacency matrix **A** is essentially a |V| x |V| symmetric square matrix where the element *aij* is equal to one if atoms *i* and *j* are adjacent, zero otherwise:



* Atoms are adjacent if there exists a chemical bond connecting them.
* Matrix-Based Representation: - Adjacency matrix is just one of the possible matrices that can encode information about a molecular graph.
* Graph-theoretical matrices can be classified as vertex matrices, edge matrices, or incidence matrices. Vertex matrices are those matrices whose rows and columns refer to graph vertices, the atoms, and each element of the matrix encodes a property associated to a pair of vertices.
* Most of the graph-theoretical matrices are calculated on a graph that does not include the hydrogen atoms, the so-called H-depleted molecular graph; however, they can be also derived from the H-filled molecular graph, if hydrogens are required to better represent the molecular structure.
* Vertex matrices are more used to derive molecular descriptors than edge and incidence matrices. The most important are the adjacency matrix and the topological distance matrix.
* One of the first molecular descriptors derived from the adjacency matrix **A** is the Lovasz-Pelikan index, which is defined as the leading eigenvalue of the adjacency matrix:



* The topological distance matrix **D** is a symmetric square |V| x |V| matrix whose element *dij* is equal to the topological distance vertices *i* and *j*:



* Here is the number of edges along the shortest path between *i* and *j*.
* The Wiener index (*W*), which is defined as the half-sum of all the elements *dij* of the distance matrix:



* The Laplacian matrix **L**, also known as Kirchhoff matrix, is a symmetric square |V| x |V| matrix defined as an augmented matrix.
* The Laplacian matrix is defined as the difference between the vertex degree matrix and the adjacency matrix.
* The vertex degree matrix is a diagonal matrix whose diagonal elements are the vertex degrees, the vertex degree being the number of non-H atoms adjacent to atom *i*:



* Vertex degrees can be calculated as the row sums of the adjacency matrix **A:**



* The Laplacian spectrum is the set of eigenvalues derived from the diagonalization of the Laplacian matrix.
* Quasi-Wiener index (*W\**) is calculated as the product of the number |V| of vertices of the graph and the sum of the reciprocal non-zero eigenvalues of the Laplacian matrix:



* Here |V|-1 is the number of positive eigenvalues and the *i*-th eigenvalue of the Laplacian matrix.
* The detour matrix , also known as maximum path matrix, is a symmetric vertex matrix defined as:



* Here is the number of edges along the longest path between vertices *i* and *j*, that is, the detour distance.
* Wiener-type defined as the half-sum of the detour distances between every pair of vertices included in the molecular graph:



* Reciprocal matrices are obtained by raising to different powers the elements of the considered matrix. A typical example of these matrices is the reciprocal distance matrix , also known as Harary matrix **H**:



* The Harary index *H* is calculated from the reciprocal distance matrix as a Wiener-type index:



* The edge weights are bond properties like the conventional bond order, the bond length, some combination of the properties of the connected atoms, or any other property that can be associated to the considered bond.
* The atom connectivity matrix **C** is the most known example of weighted matrices. This is a weighted adjacency matrix whose elements *cij* are defined as follows:



* Where is the conventional bond order.
* Another example of weighted matrices is the matrix that is defined as follows:



* Here is the vertex degree of the two connected atoms.
* Randié connectivity index () is among the first proposed topological indices; it can be calculated from the matrix as a Wiener-type index:



* Here are the adjacency matrix elements, which equal 1 for pairs of connected atoms, 0 otherwise.
* Burden matrix **B** is another weighted matrix, which provides the well-known Burden eigenvalue descriptors. Burden matrix is defined as follows:



* The most common edges matrices are edge adjacency matrix. This is a symmetric matrix whose dimension is |E| x |E|, which collects information about the connectivities of molecular bonds. Formally, it is defined as:



* To calculate the vertex degree , the edge degree can be calculated from the edge adjacency matrix as:



* Edge adjacency matrix has been used to derive different graph-theoretical invariants, such as the edge connectivity index defined as:



* Here are the elements of the edge adjacency matrix.
* Generalized Topological Indices: - Topological indices are single invariants derived from a graph-based representation of a molecular structure.
* Topological indices are usually calculated on an H-depleted molecular graph, and depending on their definition, they are responsive to one or more structural graph, and depending on their definition, they are responsive to one or more structural features of the molecule, like size, shape, symmetry, and ring system characteristic.
* Local vertex invariants are those numerical quantities of graph vertices used to characterize specific properties of the molecule’s atoms.
* Typical local invariants are the vertex degree , the valence vertex degree, and the bond vertex degree .
* Local invariants can be calculated from any graph-theoretical matrix. One of the most common approaches is to sum the values of a generic symmetric graph-theoretical matrix along a row:



* Here **M** is the considered matrix of dimension A x A.
* By applying this operator to the adjacency matrix, **A** results in the calculation of the most common local vertex invariant, the vertex degree :



* In this research the Wiener index is defined as the sum of the topological distances between all the atom pairs, but it is commonly calculated as the half-sum of the distance matrix entries.
* This allows the evaluation of the behavior of these indices when applied on matrices including different types of information.
* Let **M** be a generic A x A symmetric graph-theoretical matrix, then a collection of generalized topological indices can be defined.
* The formula *D*1 is a generalization of the Wiener index:



* In this, two parameters have been added to the original formula, a scale parameter and a power parameter . If the matrix **M** coincides with the distance matrix (**D**), = ½ and = 1, then the formula *D*1 corresponds to the Wiender index.
* If matrix **M** is the distance matrix (**D**), = ½ and = -1, then the Harary index is obtained.
* An additional parameter can be introduced in formula *D*2 to restrict the summation to adjacent elements, being the elements of the adjacency matrix:



* Based on generic local vertex invariants , a general formula for the calculation of generalized topological indices is defined as the following:



* The Balaban index *J* is obtained setting = |E|/2(*c*+1), where *r* is the cyclomatic number, and = -1/2 and considering the as the sum of all entries of the *i*-th row of the distance matrix **D**.
* The further generalization of the formula *D*3 be applied on any type of graph fragments and can be written as:



* Here, is a generic local vertex invariant, *K* is the total number of selected graph fragments, and is the number of vertices in the *k*-th fragment.
* Kier and Hall proposed a general scheme to calculate of Kier-Hall connectivity indices based on the Randić index formula but considering larger fragments than bonds:



* Here is the vertex degree of the *i*-th atom.
* Kier and Hall connectivity indices are calculated on all the subgraphs, of the predefined order *m*, constituting the molecule.
* Zero-order subgraphs (*m*=*0*) are all the atoms in the molecule; it means that the summation goes over all vertex degrees, *K* is equal to the number of atoms, and = 1.
* 1-order subgraphs (*m = 1*) are all pairs of adjacent atoms, that is, all paths of length on (i.e., bonds), and for *m* = 1 the Randić-connectivity index is obtained, with *K* equal to the number of edges |E| and equal to 2.
* Second-order connectivity indices (*m = 2*) are calculated on all paths of order 2, *K* is equal to the number pf path of length 2 in the graph, and is equal to 3 since any path of length two involves three atoms.
* Autocorrelation Descriptors: - autocorrelation descriptors denote a heterogeneous class of molecular descriptors whose general mathematical formalization is the following:



* Here is a generic local invariant: and are a scaling and a power parameter, respectively; and is Kronecker delta function equal to one for pairs of substructure centers at topological distance dij = k and zero otherwise; A is the number of substructure centers that typically are the molecule atoms.
* The most known autocorrelation descriptors are the Moreau-Broto, Moran and, Geary autocorrelation descriptors.
* Moreau and Broto applied for the first time the spatial autocorrelation function to a molecular graph to measure the distribution of atomic properties on the molecular graph.
* It can be derived from the general equation *D*5 setting = ½ and = 1 and substituting the generic local vertex invariant with any atomic property as:



* Here V is the set of vertices of the H-filled molecular graph.
* Moreau-Broto autocorrelation calculated for path length zero is defined as the sum of the squares of the atomic properties:



* The weight *w* can be defined arbitrarily.
* In the Dragon software unlike Moreau-Broto autocorrelation descriptors (ATS), they explicitly take into account the mean and the standard deviation of the properties used as the atom-weighting scheme.
* Moran autocorrelation descriptors (MATS) have been defined as:



* Here, wi is any atomic property, is the property mean on the molecule, and is the sum of the Kronecker deltas, that is, the number of atom pairs at distance equal to *k*.
* Values between 0 and 1 reveal positive autocorrelation, while values larger than 1 are associated to negative spatial autocorrelation.
* Atom-type autocorrelation descriptors are discrete descriptors, while Moreau-Broto, Moran, and Geary descriptors are continuous functions.
* Atom-type autocorrelation descriptors count the occurrences of atom pairs, where atom pairs are defined as any pair of atoms having particular features at a predefined topological distance *k*.



* Here, *u* and *v* represent two different atom types and (*i;u*), (*j;v*), and (*dij;k*) are three Kronecker delta functions equal to one if atom *i* is of type *u*, atom *j* is of type *v,* and the topological distance *dij* is equal to *k*, respectively, zero otherwise.
* Carhart descriptors are atom pair descriptors where atom types are defined using the chemical element of atoms, their connectivity calculated on the H-depleted molecular graph and the number of electrons.
* Another type of well-known atom pair descriptors are the CATS (Chemically Advanced Template Search) descriptors.
* The atom-type definition in CATS is related to the concept of “potential pharmacophore points” (PPP), where a PPP is a generalized atom type defined considering some physicochemical aspects.
* Five PPP have been originally proposed for the calculation of CATS: hydrogen-bond donor (D), hydrogen-bond acceptor (A), positive (P), negative (N), and lipophilic (L).
* The relative scaling, using the sum of the occurrences of the PPP, highlights rare PPP considering them as possible important contributors of the pharmacophoric aspect of the molecules with respect to abundant PPP pairs like the lipophilic atom types.
* Three-dimensional autocorrelation descriptors are based on Euclidean interatomic distances, while the property is still defined as an atomic property.
* CATS3D have been defined using 20 distance bins from o to 20 Å in steps of 1 Å.
* The formula for these descriptors is the following:



* Here *k* is the considered bin, *u* and *v* represent two different atom types, and (*i;u*), (*j;v*), and (*rij;k*) are three Kronecker delta functions equal to one respectively if atom *i* is of type *u*, atom *j* is of type *v,* and their interatomic distance *rij* is comprised in the *k*-th bin, zero otherwise.
* Geometrical Descriptors: - Geometrical descriptors are those molecular descriptors calculated taking into account the three-dimensional structure of the molecule, i.e., the position of the atoms in the three-dimensional space.
* Their application is bound to those cases for which a specific conformation is selected or diversity among conformations has to be evaluated.
* The geometrical distance matrix **G** is the common representation of the distance relationships between all of the atom pairs in the 3D space; this is a symmetric |V| x |V| matrix whose entries *rij* are the Euclidean distances between every pair of atoms *i* and *j* in the molecule:



* The 3D-Wiener index () is a topographic index calculated by analogy with the Wiener index from the geometrical distance matrix as:



* A local vertex invariant can be calculated from the geometrical distance matrix is the geometric distance degree, which is calculated as the row sum of the geometrical distance matrix as:



* Here *rij* is the Euclidean distance between atoms *i* and *j*.
* The 3D connectivity indices () have been defined substituting the vertex degree with the geometric distance degree :



* Here *K* is the number of *m*-th order subgraph constituted by vertices.
* Two additional examples of descriptors calculated using the geometrical distance matrix as a basis are the Euclidean connectivity () index:



* And the 3D Balaban index:



* Here *c* is the cyclomatic number, that is, the number of independent cycles in the structure.
* An example of matrix combination is the distance/distance matrix **G/D**, which is a |V| x |V| symmetric vertex matrix whose values are defined as the ratio between the geometrical distance and the topological distance:



* Here *rij* and *dij* are the Euclidean and the topological distance between atoms *i* and *j*, respectively.
* Two most commonly known classes of three-dimensional descriptors are the WHIM (Weighted Holistic Invariant Molecular) descriptors and the GETAWAY (Geometry, Topology, and Atom-Weights Assembly) descriptors.
* WHIM descriptors encode three-dimensional information on size, shape, symmetry, and atomic property distribution.
* WHIM descriptors are invariant to roto-translation: they are invariant to translation as a result of the centering of the atomic coordinates and to rotation as long as the principal axes are unique.
* The weighted covariance matrix for a generic atomic weighting scheme *w* is defined as:



* Here the summations run over the molecule atoms, is the weighted covariance between the two sets of atomic coordinates *j* and *k* (*j, k = x, y, z*), *wi* is the atomic property of atom *i*, and , , and represent the coordinates of the *i*-th and *j*-th atoms and the corresponding average value, respectively.
* The three eigenvalues of the weighted covariance matrix play a crucial role in the WHIM descriptor calculation.
* Directional WHIM descriptors consider the three eigenvalues separately providing information on the molecular size, shape, density, and symmetry along the principal axis, while global WHIM descriptors consider the relationships among the eigenvalues to describe molecular properties along the three principal directions in the molecule.
* GETAWAY descriptors are three-dimensional descriptors that encode information about the influence that each atom has in determining the whole shape in the molecule and evaluate the interactions among atoms with respect to their geometrical position in the three-dimensional space.
* The first matrix is called molecular influence matrix **H**, which is an A x A symmetric matrix, defined as:



* Here, **M** is the molecular matrix of the centered Cartesian coordinates (x, y, z) calculated on the H-filled three-dimensional molecular graph for a defined conformation.
* Influence matrix **H** is invariant to rotation of the molecular coordinates, granting that GETAWAY descriptors are independent of any alignment.
* Elements of the influence matrix split with respect to their meaning, in two sets.
* The former comprises the diagonal elements of **H**, denoted as *hi* and called leverages, whose range is from 0 to 1.
* The latter includes the off-diagonal elements *hij* that reflect the capability of the *j*-th atom to interact with the *i*-th atom.
* The second matrix to be considered for GETAWAY calculation is the influence/distance matrix **R**, an A x A symmetric matrix determined combining the elements of the influence matrix **H** with the elements of the geometrical distance matrix **G**:



* Here *hi* and *hj* are the leverages of atoms *i* and *j*, respectively, and *rij* is the Euclidean distance between them.
* Structural Keys and Molecular Fingerprints: - structural keys and molecular fingerprints try to describe a chemical capturing different local aspect of a molecular structure. Specifically, these approaches identify a set of structural fragments participating in the molecule composition.
* Structural keys are a predefined set of structural features, atom pairs, functional groups, ring systems, and atom-centered fragments that have been prepared in order to be able to discriminate among molecules and eventually be able to highlight chemical properties of a molecule.
* The most known structural keys are the two sets of MACCS keys comprising 960 and 166 structural features, respectively, and the PubChem structural keys, which include 881 bits reflecting the presence or absence of 881 structural features.
* The molecular fingerprints are identified exploding the molecular structure in all possible substructure patterns following a set of rules.
* Substructure patterns can be identified increasing or lowering the discriminatory power just defining the rules to be used to recognize and differentiate the patterns.
* Two major types of patterns can be identified: paths and atom centered, also known as circular patterns. Then two classes of fingerprints can be calculated, path and extended connectivity fingerprint, also known as Morgan or circular fingerprint.
* The first parameter to be set once the type of fingerprint has been defined is the maximum path length or the maximum radius of the patterns to be identified. Then, additional options can be set in order to increase or decrease the discrimination power of substructures.
* Typically, molecular fingerprint is a Boolean vector coding the presence or absence of structural features.
* To obtain a fixed-length vector of molecular fragments, the Boolean vector constituting the fingerprint is processed using hash functions.
* A hash function is any function that can be used to reduce a variable-length Boolean vector to a fixed-length one.
* Each identified pattern is used as pseudorandom number generator seed, and the output of the hash function is a set of bits.
* Fingerprint hashing is introduced when it is not possible to assign a particular bit to each pattern, as for structural keys, due to the huge number of possible identifiable patterns.
* Hashing functions enable the definition of a fixed-length vector but introduce the so-called bit collision; this means that, under a predefined set of rules, a specific fragment will always be associated to a defined set of bits in the fingerprint, but two different fragments may share one or more bits among their bit sets.
* Fingerprints encode an almost exhaustive set of patterns with respect to structural keys, resulting in more detailed description of a molecular structure in almost all situations.
* Future Perspectives: - it is quite recent, and now able to provide solid answers to real problems. Molecular descriptors theory is facing new challenges requiring both chemical and mathematical skills. The description of structures not considered before, like disconnected structures and nanomaterials, is one of these challenges.
* Molecular Descriptors for Disconnected Structures: - one of the challenges that molecular modeling is facing is the capability to be applied not only on small organic molecules but also on different kinds of chemicals, like disconnected structures.
* Besides data availability, one of the relevant issues concerning modeling of disconnected structures in the representation of these structures through appropriate molecular descriptors.
* Considering, the calculation algorithm, the classical descriptors can be divided into two classes. The former comprises all the descriptors whose mathematical definition does not require that the molecule is full connected; the latter includes those descriptors that can be calculated only on a fully connected graph.
* Molecular weight is a molecular descriptor whose meaning is preserved on disconnected structure; a cyclomatic number is another descriptor whose definition comprises the possibility to be calculated on disconnected structures.
* Functional groups, atom-centered fragments, fingerprints, and structural keys are additive descriptors whose interpretation is identical for fully connected and non-fully connected structures.
* The first-class descriptors can be directly calculated on disconnected structures, the second class requires a deeper evaluation on how these descriptors can be calculated on disconnected structures.
* Molecular Descriptors for Nanostructures: - Nanomaterials are usually defined as those materials of which a single unit size is between 1 and 100 nanometers. QSAR applied to nanomaterials, named nano-QSAR or QNAR (quantitative nanostructure-activity relationships), is an emerging research field whose preliminary steps have been moved.
* Due to the lack of theoretical descriptors, currently most of the molecular descriptors used to provide nanostructure description are experimental descriptors.
* Till now the data availability is one of the major bottlenecks for the advancement in nano-QSAR studies.
* Conclusions: - the molecular graph is the most common representation of the molecular structure, and from it, applying a finite set of operators, a huge number of descriptors can be derived.
* Most of them have clear chemical interpretation but others rather seem to be the outcome of some mathematical trick.
* In short, when dealing with molecular descriptors, one should keep in mind that molecular descriptors are numbers that can be easily generated through different combinations of structure representations, atom/bond weighting schemes, and mathematical functions, but in order to produce useful descriptors, some basic rules should be fulfilled not only related to their mathematical definition but also to their chemical interpretation and the correlation with one or more experimental property.