E

➢ **EA indices** ≡ *Extended Adjacency matrix indices* → spectral indices
➢ **EAmax index** → spectral indices (⊙ extended adjacency matrix indices)
➢ **EAΣ index** → spectral indices (⊙ extended adjacency matrix indices)
➢ **eccentric** → distance matrix
➢ **eccentric adjacency index** → eccentricity-based Madan indices (⊙ Table E1)
➢ **eccentric adjacency topochemical indices** → eccentricity-based Madan indices (⊙ Table E1)

■ **eccentricity-based Madan indices**

These are a series of → *graph invariants* calculated on the → *H-depleted molecular graph* and defined as different combinations of → *atom eccentricity* η, which is the maximum topological distance from an atom, and other → *LOcal Vertex Invariants*. Eccentricity-based Madan indices are listed in Table E1 where index names, formulas, and bibliographic references are given.

These consist of two sets of indices: one is topological (ID 1-7) as it accounts only for features of a simple graph; the other set (ID 8-12) was defined topochemical as it was derived from vertex-weighted graphs to account for heteroatoms.

Together with atom eccentricity η, local invariants used to derive these molecular descriptors are → *vertex degree* δ_i, which is the number of vertices bonded to the ith vertex; → *distance degree* σ_i, which is the sum of the topological distances in the graph from vertex v_i; Morgan's → *extended connectivity* of first order, EC_i^1, obtained by summing up the vertex degrees of all vertices bonded to vertex v_i; → *Madan chemical degree*, δ_i^c, calculated by summing up the relative atomic weights of the adjacent vertices; → *chemical atom eccentricity*, η_i^c, which is the maximum distance weighted by relative atomic weights from the considered vertex; → *chemical extended connectivity*, EC_i^{1c}, obtained by summing up the Madan chemical degrees of all first neighbor vertices. M_i, used for the calculation of the augmented eccentric connectivity index (ID-7), is the product of the degrees of all the vertices adjacent to the vertex v_i [Bajaj, Sambi *et al.*, 2006a]:

$$M_i = \prod_{j=1}^{A} (\delta_j)^{a_{ij}}$$

where the exponent a_{ij} takes value 1 only for pairs of adjacent vertices, zero otherwise, thus giving a contribution of 1 in the product. This local invariant is a modification of the extended connectivity and thus may be called **augmented connectivity**.

Table E1 Definitions and bibliographic references of the eccentricity-based Madan indices.

ID	Descriptor name	Formula	Reference
1	Eccentric connectivity index	$\xi^c = \sum_{i=1}^{A} \eta_i \cdot \delta_i$	[Sharma, Goswami *et al.*, 1997; Sardana and Madan, 2002c]
2	Eccentric distance sum	$\xi^{DS} = \sum_{i=1}^{A} \eta_i \cdot \sigma_i$	[Gupta, Singh *et al.*, 2002]
3	Adjacent eccentric distance sum index	$\xi^{SV} = \sum_{i=1}^{A} \dfrac{\eta_i \cdot \sigma_i}{\delta_i}$	[Sardana and Madan, 2002b]
4	Connective eccentricity index	$C^\xi = \sum_{i=1}^{A} \dfrac{\delta_i}{\eta_i}$	[Gupta, Singh *et al.*, 2000]
5	Eccentric adjacency index	$\xi^A = \sum_{i=1}^{A} \dfrac{EC_i^1}{\eta_i}$	[Gupta, Singh *et al.*, 2001b]
6	Superadjacency index	$^A\xi = \sum_{i=1}^{A} \dfrac{\delta_i \cdot EC_i^1}{\eta_i}$	[Bajaj, Sambi *et al.*, 2004b]
7	Augmented eccentric connectivity index	$^{Ac}\xi = \sum_{i=1}^{A} \dfrac{M_i}{\eta_i}$	[Bajaj, Sambi *et al.*, 2006a]
8	Superadjacency topochemical index	$^{Ac}\xi = \sum_{i=1}^{A} \dfrac{\delta_i^c \cdot EC_i^{1c}}{\eta_i^c}$	[Bajaj, Sambi *et al.*, 2004b]
9	Eccentric connectivity topochemical index	$\xi_c^c = \sum_{i=1}^{A} \eta_i^c \cdot \delta_i^c$	[Kumar, Sardana *et al.*, 2004]
10	Eccentric adjacency topochemical index (1)	$\xi_{1C}^A = \sum_{i=1}^{A} \dfrac{EC_i^{1c}}{\eta_i}$	[Gupta, Singh *et al.*, 2003]
11	Eccentric adjacency topochemical index (2)	$\xi_{2C}^A = \sum_{i=1}^{A} \dfrac{EC_i^1}{\eta_i^c}$	[Gupta, Singh *et al.*, 2003]
12	Eccentric adjacency topochemical index (3)	$\xi_{3C}^A = \sum_{i=1}^{A} \dfrac{EC_i^{1c}}{\eta_i^c}$	[Gupta, Singh *et al.*, 2003]

Topochemical indices are calculated from three graph theoretical matrices: (1) a → *chemical adjacency matrix* obtained by substituting row elements of the adjacency matrix, corresponding to bonded atoms, with relative atomic weights of the bonded atoms; (2) an → *atomic weight-weighted distance matrix* obtained by substituting topological distances with the sum of relative atomic weights of all the vertices involved in the shortest path; and (3) an → *additive chemical adjacency matrix* obtained by modifying the → *additive adjacency matrix*, replacing the vertex degree by the → *Madan chemical degree*.

Example E1

Eccentricity-based Madan indices for 2-pentanol.
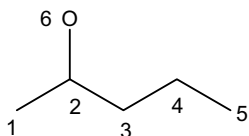
Vertex degrees δ

Atom	1	2	3	4	5	6
δ	1	3	2	2	1	1

Distance matrix

Atom	1	2	3	4	5	6	σ_i
1	0	1	2	3	4	2	12
2	1	0	1	2	3	1	8
3	2	1	0	1	2	2	8
4	3	2	1	0	1	3	10
5	4	3	2	1	0	4	14
6	2	1	2	3	4	0	12

Atom	η_i
1	4
2	3
3	2
4	3
5	4
6	4

Additive adjacency matrix

Atom	1	2	3	4	5	6	EC_i^1
1	0	3	0	0	0	0	3
2	1	0	2	0	0	1	4
3	0	3	0	2	0	0	5
4	0	0	2	0	1	0	3
5	0	0	0	2	0	0	2
6	0	3	0	0	0	0	3

Eccentric connectivity index:

$$\xi^c = \sum_{i=1}^{6} \eta_i \cdot \delta_i = 4 \cdot 1 + 3 \cdot 3 + 2 \cdot 2 + 3 \cdot 2 + 4 \cdot 1 + 4 \cdot 1 = 31$$

Eccentric distance sum:

$$\xi^{DS} = \sum_{i=1}^{6} \eta_i \cdot \sigma_i = 4 \cdot 12 + 3 \cdot 8 + 2 \cdot 8 + 3 \cdot 10 + 4 \cdot 14 + 4 \cdot 12 = 222$$

Adjacency eccentric distance sum index:

$$\xi^{SV} = \sum_{i=1}^{6} \frac{\eta_i \cdot \sigma_i}{\delta_i} = \frac{4 \cdot 12}{1} + \frac{3 \cdot 8}{3} + \frac{2 \cdot 8}{2} + \frac{3 \cdot 10}{2} + \frac{4 \cdot 14}{1} + \frac{4 \cdot 12}{1} = 183$$

Connective eccentricity index:

$$C^{\xi} = \sum_{i=1}^{6} \frac{\delta_i}{\eta_i} = \frac{1}{4} + \frac{3}{3} + \frac{2}{2} + \frac{2}{3} + \frac{1}{4} + \frac{1}{4} = 3.417$$

Eccentric adjacency index:

$$\xi^A = \sum_{i=1}^{6} \frac{EC_i^1}{\eta_i} = \frac{3}{4} + \frac{4}{3} + \frac{5}{2} + \frac{3}{3} + \frac{2}{4} + \frac{3}{4} = 6.833$$

Superadjacency index:

$$\int^A = \sum_{i=1}^{6} \frac{\delta_i \cdot EC_i^1}{\eta_i} = \frac{1 \cdot 3}{4} + \frac{3 \cdot 4}{3} + \frac{2 \cdot 5}{2} + \frac{2 \cdot 3}{3} + \frac{1 \cdot 2}{4} + \frac{1 \cdot 3}{4} = 13$$

The multiplicative connectivities of the vertices are

$$M_1 = \delta_2 = 3, \qquad M_2 = \delta_1 \cdot \delta_3 \cdot \delta_6 = 2, \qquad M_3 = \delta_2 \cdot \delta_4 = 6,$$
$$M_4 = \delta_3 \cdot \delta_5 = 2, \qquad M_5 = \delta_4 = 2, \qquad M_6 = \delta_2 = 3$$

Augmented eccentric connectivity index:

$$^A\xi^c = \sum_{i=1}^{6} \frac{M_i}{\eta_i} = \frac{3}{4} + \frac{2}{3} + \frac{6}{2} + \frac{2}{3} + \frac{2}{4} + \frac{3}{4} = 6.333$$

For the calculation of the topochemical indices the relative atomic weights ($m_1 = m_2 = m_3$ $m_4 = m_5 = 1$, and $m_6 = 1.332$) and the following graph-theoretical matrices were used:

Chemical adjacency matrix

Atom	1	2	3	4	5	6	δ_i^c
1	0	1	0	0	0	0	1
2	1	0	1	0	0	1.332	3.332
3	0	1	0	1	0	0	2
4	0	0	1	0	1	0	2
5	0	0	0	1	0	0	1
6	0	1	0	0	0	0	1

Additive chemical adjacency matrix

Atom	1	2	3	4	5	6	EC_i^{1c}
1	0	3.332	0	0	0	0	3.332
2	1	0	2	0	0	1	4
3	0	3.332	0	2	0	0	5.332
4	0	0	2	0	1	0	3
5	0	0	0	2	0	0	2
6	0	3.332	0	0	0	0	3.332

Atomic weight-weighted distance matrix

Atom	1	2	3	4	5	6	η_i^c
1	0	1	2	3	4	2.332	4
2	1	0	1	2	3	1.332	3
3	2	1	0	1	2	2.332	2.332
4	3	2	1	0	1	3.332	3.332
5	4	3	2	1	0	4.332	4.332
6	2	1	2	3	4	0	4

Superadjacency topochemical index:

$$\int^{A_c} = \sum_{i=1}^{6} \frac{\delta_i^c \times EC_i^{1c}}{\eta_i^c} = \frac{1 \times 3.332}{4} + \frac{3.332 \times 4}{3} + \frac{2 \times 5.332}{2.332} + \frac{2 \times 3}{3.332}$$
$$+ \frac{1 \times 2}{4.332} + \frac{1 \times 3.332}{4} = 12.944$$

Eccentric connectivity topochemical index:

$$\xi_c^c = \sum_{i=1}^{6} \eta_i^c \cdot \delta_i^c = 1 \times 4 + 3.332 \times 3 + 2 \times 2.332 + 2 \times 3.332 + 1 \times 4.332 + 1 \times 4 = 33.656$$

Eccentric adjacency topochemical indices:

$$\xi_{1C}^A = \sum_{i=1}^{6} \frac{EC_i^{1c}}{\eta_i} = \frac{3.332}{4} + \frac{4}{3} + \frac{5.332}{2} + \frac{3}{3} + \frac{2}{4} + \frac{3.332}{4} = 7.165$$

$$\xi_{2C}^A = \sum_{i=1}^{6} \frac{EC_i^1}{\eta_i^c} = \frac{3}{4} + \frac{4}{3} + \frac{5}{2.332} + \frac{3}{3.332} + \frac{2}{4.332} + \frac{3}{4} = 6.339$$

$$\xi_{3C}^A = \sum_{i=1}^{6} \frac{EC_i^{1c}}{\eta_i^c} = \frac{3.332}{4} + \frac{4}{3} + \frac{5.332}{2.332} + \frac{3}{3.332} + \frac{2}{4.332} + \frac{3.332}{4} = 6.648$$

📖 [Sardana and Madan, 2003; Kumar and Madan, 2004]

- ➤ **eccentric connectivity index** → eccentricity-based Madan indices (⊙ Table E1)
- ➤ **eccentric connectivity topochemical index** → eccentricity-based Madan indices (⊙ Table E1)
- ➤ **eccentric distance sum** → eccentricity-based Madan indices (⊙ Table E1)
- ➤ **eccentricity** → distance matrix
- ➤ **eccentricity** ≡ *geometrical eccentricity*
- ➤ **ECFC fingerprints** → substructure descriptors (⊙ fingerprints)
- ➤ **ECFP fingerprints** ≡ *Extended Connectivity FingerPrints* → substructure descriptors (⊙ fingerprints)
- ➤ **EC method** ≡ *Electron-Conformational method* → Electronic-Topological method
- ➤ **ECN index** → electronegativity-based connectivity indices
- ➤ **ECP index** → electronegativity-based connectivity indices

◼ **edge adjacency matrix (E)**
Derived from the → *molecular graph G*, the edge adjacency matrix, denoted by **E**, or more formally as **^{E}A**, also called **bond matrix**, encodes information about the connectivity between graph edges:

$$[^E\mathbf{A}]_{ij} \equiv [\mathbf{E}]_{ij} = \begin{cases} 1 & \text{if } (i,j) \text{ are adjacent bonds} \\ 0 & \text{otherwise} \end{cases}$$

It is a square symmetric matrix of dimension $B \times B$, where B is the number of bonds, and is usually derived from a → *H-depleted molecular graph* [Bonchev, 1983]. It is to be noted that the edge adjacency matrix of a graph G is equal to the → *adjacency matrix* of the → *line graph* of G [Gutman and Estrada, 1996].

The entries $[\mathbf{E}]_{ij}$ of the matrix are equal to one if edges e_i and e_j are adjacent (the two edges thus forming a → *path* of length two) and zero otherwise. For multigraphs, the edge adjacency matrix can be augmented by a row and a column for each multiple edge.

The **edge degree** ε_i provides the simplest information related to the bond considered and is calculated from the edge adjacency matrix as follows [Bonchev, 1983]:

$$\varepsilon_i \equiv VS_i(\mathbf{E}) = \sum_{j=1}^{B} [\mathbf{E}]_{ij}$$

where VS_i is the → *row sum operator*.

The edge degree ε is related to the $\rightarrow$ *vertex degree* δ by the following relationship:

$$\varepsilon_i = \delta_{i(1)} + \delta_{i(2)} - 2$$

where $i(1)$ and $i(2)$ refer to the two vertices incident to the ith edge.

The number of edges whose same edge degree is equal to g is called the **edge degree count** ^gF_E; therefore, the vector

$$\{^1F_E;^2F_E;\ldots;^6F_E\}$$

can be associated with each graph G, provided the maximum edge degree is equal to 6.

Related to the previous definition is the **edge type count** $ne_{gg'}$, defined as the number of edges with the same vertex degree of the incident vertices, where g and g' are the degree values of the incident vertices.

The **total edge adjacency index** A_E, also known as **Platt number**, F [Platt, 1947, 1952], is the sum over all entries of the edge adjacency matrix:

$$A_E \equiv F = \sum_{i=1}^{B}\sum_{j=1}^{B}[\mathbf{E}]_{ij} = \sum_{i=1}^{B}\varepsilon_i = 2 \cdot N_2$$

where N_2 is the number of graph connections.

Two adjacent edges constitute a second-order path and this subgraph is called **connection** (or **link**). The **connection number** N_2, also known as **Gordon–Scantlebury index** (N_{GS}) [Gordon and Scantlebury, 1964], is the simplest graph invariant obtained from the edge adjacency matrix that considers both vertices and edges and is calculated as

$$N_2 \equiv N_{GS} \equiv {}^2P = A_E/2$$

where A_E is the total edge adjacency index and ^2P is the second-order path count. Note that the connection number also coincides with the $\rightarrow$ *Bertz branching index*.

The number of connections N_2 of a molecular graph is related to the $\rightarrow$ *first Zagreb index* M_1 and the $\rightarrow$ *quadratic index* Q by the following relationships:

$$N_2 = M_1/2 - A + 1 = Q + A - 2$$

where A is the number of atoms.

When multiple bonds are considered, the connection number can be calculated on multi-graphs by using more general approaches. Given an H-depleted molecular graph G, the number of connections is equal to the number of edges in the line graph of G. Moreover, the connection number can also be calculated as the sum of vertex contributions, taking into account that each connection intersects at a particular vertex and is therefore simply a function of the number of hydrogens on that atom:

$$N_2 = \frac{1}{2} \cdot \sum_{i=1}^{A}(4-h_i) \cdot (3-h_i) - DB - 3 \cdot TB$$

where h_i is the number of hydrogen atoms attached to the ith atom and DB and TB denote the number of double and triple bonds in the molecule, respectively; the sum runs over all nonhydrogen atoms in the molecule [Hendrickson, Huang *et al.*, 1987].

As the number of connections is sensitive to different features of molecular structure such as size, branching, cyclicity, and multiple bonds, it was used by Bertz to define its → *molecular complexity* index.

Example E2

Line graph of 2-methylpentane and its edge adjacency matrix. ε_i is the edge degree and ε the edge connectivity index.

edge	a	b	c	d	e	ε_i
a	0	1	0	0	1	2
b	1	0	1	0	1	3
c	0	1	0	1	0	2
d	0	0	1	0	0	1
e	1	1	0	0	0	2
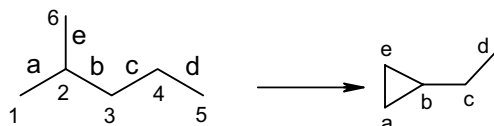
$$A_E = \sum_{i=1}^{5}\sum_{j=1}^{5}[\mathbf{E}]_{ij} = \sum_{i=1}^{5}\varepsilon_i = 2 \cdot N_2 = 10 \qquad \mathbf{E} =$$

$$\varepsilon = (2 \cdot 3)^{-1/2} + (3 \cdot 2)^{-1/2} + (2 \cdot 1)^{-1/2} + (2 \cdot 2)^{-1/2} + (3 \cdot 2)^{-1/2} = 2.432$$

From the edge adjacency matrix, a graph-theoretical invariant analogous to the → *Randić connectivity index* was derived by Estrada [Estrada, 1995a; Cash, 1995b]; it is called **edge connectivity index** (or **bond connectivity index**), denoted by ε, and defined as

$$\varepsilon = \sum_k (\varepsilon_i \cdot \varepsilon_j)_k^{-1/2}$$

where k runs over all connections N_2, and ε_i and ε_j are the edge degrees of the two edges in the connection. It coincides with the Randić connectivity index χ of the line graph of G.

The **extended edge connectivity indices** (or **bond connectivity indices**) were defined as a generalization of the edge connectivity index in analogy to the → *Kier–Hall connectivity indices*:

$$^m\varepsilon_t = \sum_{k=1}^{K} \left(\prod_i \varepsilon_i \right)_k^{-1/2}$$

where k runs over all of the mth order subgraphs, m is the number of edges in the subgraph, and K is the total number of mth order subgraphs; the edge degrees of all the edges in the subgraph are considered. The subscript "t" for the connectivity indices refers to the type of → *molecular subgraph* and is "ch" for chain or ring, "pc" for path-cluster, "c" for cluster, and "p" for path (this can also be omitted) [Estrada, Guevara *et al.*, 1998a; Estrada and Rodriguez, 1999; Estrada, 1999b]. Some mathematical relationships between the extended edge connectivity indices and → *line graph connectivity indices* were found.

The **spectral moments of the edge adjacency matrix E** were defined [Estrada, 1996, 1997, 1998b] as

$$\mu^k = tr(\mathbf{E}^k)$$

where k is the power of the edge adjacency matrix and tr the $\rightarrow$ *trace*, that is, the sum of the diagonal elements. The zero-order spectral moment μ^0 corresponds to the number of edges in the graph, that is, the trace of the resulting B-dimensional identity matrix.

Since the kth-order spectral moment μ^k corresponds to the sum of all self-returning walks of length k in the line graph of the molecular graph G, it can be expressed as the linear combination of the $\rightarrow$ *embedding frequencies* of the molecular graph, that is, the counts of different structural fragments (subgraphs) in the graph, each fragment corresponding to a specific self-returning walk. Several relations between spectral moments and embedding frequencies were derived by Estrada [Estrada, 1998b; Marković and Gutman, 1999]. A **local spectral moment** of the edge adjacency matrix is defined as the sum of diagonal entries of the different powers of the edge adjacency matrix corresponding to a molecular fragment [Estrada and Molina, 2001b, 2001c; Estrada and Gonzalez, 2003]. It is defined as

$$\mu^k(F) = \sum_{b=1}^{B_F} [\mathbf{E}^k]_{bb}$$

where F indicates the molecular fragment considered, and the summation goes over all the bonds forming the fragment. **Bond spectral moments** are obtained when the fragment corresponds to a single bond; therefore, they simply are the single diagonal entries of the bond matrix raised to the kth power. Consequently, the total spectral moments of kth order can be expressed as the sum of the bond spectral moments of the same order.

The **weighted edge adjacency matrix** $^w\mathbf{E}$ is derived from an edge-weighted molecular graph obtained by applying any edge $\rightarrow$ *weighting scheme w*, which encodes information about each bond in the molecule. The weighting scheme can be based on quantities directly characterizing bonds, such as bond distances or bond dipoles, or quantities derived from weights of those atoms involved in each bond, such as atomic mass, atomic electronegativity, surface area contribution of polar atoms, atomic charges, and so on.

A weighted edge adjacency matrix $^w\mathbf{E}$ can be calculated as

$$[^w\mathbf{E}]_{ij} = \begin{cases} w_j & \text{if } (i,j) \text{ are adjacent bonds} \\ 0 & \text{otherwise} \end{cases}$$

where the off-diagonal elements of the matrix for a weighted graph are zero for nonadjacent edges, while, if two edges e_i and e_j are adjacent, the entry i–j is defined by the weight w_j of the jth edge and the symmetric entry j–i is defined by the weight w_i of the ith edge, thus resulting in an unsymmetrical matrix.

The **weighted edge degree** $^w\varepsilon$ is calculated by applying the $\rightarrow$ *vertex sum operator VS* to the weighted edge adjacency matrix $^w\mathbf{E}$ as

$$^w\varepsilon_i = VS_i(^w\mathbf{E}) = \sum_{j=1}^{B} [^w\mathbf{E}]_{ij} = \sum_{j=1}^{B} [\mathbf{E}]_{ij} \cdot w_j$$

where B is the number of edges in the molecular graph and w the edge-weighting scheme. The weighted edge degree of each ith edge is then the summation of the weights of all the edges adjacent to ith edge, that is, $[\mathbf{E}]_{ij} = 1$.

The **bond order-weighted edge adjacency matrix** $^\pi\mathbf{E}$ is obtained for weighted graphs whose edges are weighted by the $\rightarrow$ *bond order π* calculated by $\rightarrow$ *computational chemistry* methods on

selected molecular geometries [Estrada and Montero, 1993]:

$$[^{\pi}\mathbf{E}]_{ij} = \begin{cases} \pi_j & \text{if } (i,j) \text{ are adjacent bonds} \\ 0 & \text{otherwise} \end{cases}$$

Therefore, the ith edge degree calculated on the weighted edge adjacency matrix is the sum of the bond orders associated with all ε_i bonds adjacent to the ith edge:

$$^{\pi}\varepsilon_i = \sum_{j=1}^{B} [^{\pi}\mathbf{E}]_{ij} = \sum_{j=1}^{B} [\mathbf{E}]_{ij} \cdot \pi_j$$

The **bond order-weighted edge connectivity index** $^{\pi}\varepsilon$ is then defined as [Estrada and Ramirez, 1996]

$$^{\pi}\varepsilon = \sum_k \left(^{\pi}\varepsilon_i \cdot {}^{\pi}\varepsilon_j\right)_k^{-1/2}$$

where k runs over all connections N_2.

Characteristics of this index are its sensitivity to the presence of heteroatoms and heteroatom position in the molecule (greater values referring to central positions), and the discriminating power of conformational isomers.

By analogy with the bond order-weighted edge adjacency matrix, a **resonance-weighted edge adjacency matrix** $^{k}\mathbf{E}$ was also proposed, replacing the bond orders with parameters k_{C-X} used in the Hückel matrix and related to the resonance integral β_{C-X} of the bond between the heteroatom X and the carbon atom by the relationship:

$$\beta_{C-X} = k_{C-X} \cdot \beta_{C-C}$$

where β_{C-C} is the resonance integral of the carbon–carbon bond [Estrada, 1995b]. The literature reports several values for k_{C-X} parameters and some proposed by Estrada are listed in Table E2.

Table E2 Values of the k_{C-X} constants proposed by Estrada.

C—X bond	k_{C-X}	C—X bond	k_{C-X}
C—C	1.0	C—S	0.7
C—N	1.0	C—F	0.7
C—O	0.8	C—Cl	0.4
C=O	1.6	C—Br	0.3

The **resonance-weighted edge connectivity index** $^{k}\varepsilon$ is derived from the row sums of the above-defined matrix in the same way as the edge connectivity index ε and it is sensitive to the presence of heteroatoms and multiple bonds in the molecule.

The **electronegativity-weighted edge connectivity index**, denoted by ^{m}F, was calculated from an electronegativity-weighted edge adjacency matrix $^{\chi}\mathbf{E}$ and by the formula of the extended edge connectivity indices [Mu and Feng, 2004; Mu, Feng *et al.*, 2006]:

$$^{m}F = \sum_{k=1}^{K} \left(\prod_{i=1}^{m} {}^{\chi}\varepsilon_i\right)_k^{-1/2}$$

where k runs over all of the mth-order path subgraphs, m being the number of edges in the path subgraph, K is the total number of mth-order paths. For each path, the electronegativity-weighted edge degrees $^\chi\varepsilon_i$ of all edges involved in the path are multiplied.

To calculate this weighted edge connectivity index, the following weighting scheme is adopted:

$$w_i = \frac{\chi^h_{i(1)} + \chi^h_{i(2)}}{\chi^h_C + \chi^h_C} = \frac{\chi^h_{i(1)} + \chi^h_{i(2)}}{4.96}$$

where χ^h denotes the hybridized-dependent electronegativities (Table E3) and $i(1)$ and $i(2)$ refer to the two atoms forming the ith bond.

Table E3 Hybridized-dependent electronegativities for carbon and oxygen atoms.

Atom hybrid	C_{sp^3}	C_{sp^2}	C_{sp}	O_{sp^3}	O_{sp^2}
χ^h	2.48	2.75	3.29	4.93	5.54

By analogy with the $\rightarrow$ *augmented adjacency matrix*, the **augmented edge adjacency matrix** $^a\mathbf{E}(w)$ can be derived from an edge-weighted molecular graph, for any $\rightarrow$ *weighting scheme w* as

$$[^a\mathbf{E}(w)]_{ij} = \begin{cases} 1 & \text{if } (i,j) \text{ are adjacent bonds} \\ w_i & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

where the diagonal elements are different from zero in order to encode information about the different bonds in the molecule and edge adjacencies are codified as in the standard edge adjacency matrix.

The **bond distance-weighted edge adjacency matrix** $^a\mathbf{E}(r)$ is an augmented edge adjacency matrix based on bond lengths calculated by computational chemistry methods [Estrada, 1997]:

$$[^a\mathbf{E}(r)]_{ij} = \begin{cases} 1 & \text{if } (i,j) \text{ are adjacent bonds} \\ r_i & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

where r_i is the bond length associated to the ith edge. Bond lengths are used as weights for edges of the molecular graph, thus allowing discrimination among the different isomers, heteroatoms, conformations, and so on.

The **spectral moments of the bond distance-weighted edge adjacency matrix** were defined [Estrada, 1997; Estrada, 1998a] as

$$\mu^k(r) = tr[^a\mathbf{E}^k(r)]$$

where k is the power of the bond distance-weighted edge adjacency matrix and tr the trace of this matrix. In a similar way, spectral moments of any weighted edge adjacency matrix can be calculated by using different weighting schemes for the bonds.

A graph-theoretical approach called **TOPological Substructural MOlecular DEsign (TOPS-MODE)**, previously called **TOpological SubStructural MOlecular DEsign (TOSS-MODE)**, was

proposed to express physical and biological properties in terms of substructural features of molecules by the spectral moments of the edge adjacency matrix [Estrada, 1998b, 2008; Estrada, Peña *et al.*, 1998; Estrada and Uriarte, 2001a, 2003; Estrada, Molina *et al.*, 2001b]. The main steps to be conducted for the application of this approach to QSAR/QSPR modeling are (a) compute the spectral moments of the weighted edge adjacency matrix with the selected weighting scheme for each molecule in the data set; (b) find a QSAR/QSPR model by using any appropriate linear or nonlinear multivariate statistical technique; and (c) replace the spectral moments in the QSAR/QSPR model with their expressions in terms of the contributions of the different structural fragments of the molecule, obtaining an equation that relates the property directly with the molecular structure and thus allows → *reversible decoding*.

Applications of the TOPS-MODE approach reported in literature are [Estrada, Gutierrez *et al.*, 2000; Estrada, Uriarte *et al.*, 2000; Estrada and Peña, 2000; Estrada, Molina *et al.*, 2001a; Estrada, Vilar *et al.*, 2002; Estrada, Patlewicz *et al.*, 2003; Estrada and Gonzalez, 2003; Pérez González and Helguera, 2003; Pérez González, Gonzalez *et al.*, 2003, 2003; Estrada, Patlewicz *et al.*, 2004; Estrada, Quincoces *et al.*, 2004; Pérez González and Moldes Teran, 2004; Pérez González, Helguera Morales *et al.*, 2004; Pérez González, Helguera *et al.*, 2004; Vilar, Estrada *et al.*, 2005; Amić, Davidović-Amić *et al.*, 2007].

📖 [Estrada, Rodriguez *et al.*, 1997; Nikolić and Trinajstić, 1998, 1998; Marković, 1999; Estrada and Peña, 2000; Estrada, Uriarte *et al.*, 2000, 2003 ; Estrada, Gutierrez *et al.*, 2000; Estrada, Molina *et al.*, 2001a, 2001b; Estrada and Uriarte, 2001a; Marković, Marković *et al.*, 2001, 2001b; Estrada, Vilar *et al.*, 2002; Estrada and Gonzalez, 2003; Estrada, Patlewicz *et al.*, 2003; Marković, 2003; Pérez González and Helguera, 2003; Pérez González, Gonzalez *et al.*, 2003, 2004; Estrada, Quincoces *et al.*, 2004; Pérez González, Helguera Morales *et al.*, 2004, 2006; Pérez González and Moldes Teran, 2004; Pérez González, Helguera *et al.*, 2004; Vilar, Estrada *et al.*, 2005; Helguera Morales, Cabrera Pérez *et al.*, 2006]

➢ **edge-adjacency-plus-edge-distance matrix** → Schultz molecular topological index
➢ **edge centric indices** → centric indices
➢ **edge centric indices for multigraphs** → centric indices
➢ **edge chromatic decomposition** → chromatic decomposition
➢ **edge chromatic information index** → chromatic decomposition
➢ **edge chromatic number** → chromatic decomposition
➢ **edge-Cluj matrices** → Cluj matrices
➢ **edge complete centric index** → centric indices
➢ **edge connectivity** → connectivity indices
➢ **edge connectivity index** → edge adjacency matrix
➢ **edge-connectivity matrix** $\equiv \chi^E matrix$ → weighted matrices (⊙ weighted adjacency matrices)
➢ **edge counting** ≡ *bond number*
➢ **edge-cycle incidence matrix** → incidence matrices (⊙ cycle matrices)
➢ **edge cyclic degree** → incidence matrices (⊙ cycle matrices)
➢ **edge degree** → edge adjacency matrix
➢ **edge degree count** → edge adjacency matrix
➢ **edge degree–distance index** → Cao–Yuan indices
➢ **edge-degree Zagreb indices** → Zagreb indices
➢ **edge distance code** → edge distance matrix

> ➤ **edge distance code centric index** → centric indices
> ➤ **edge distance counts** → edge distance matrix
> ➤ **edge distance degree** → edge distance matrix
> ➤ **edge distance degree centric index** → centric indices
> ➤ **edge-distance-edge-degree matrix** → distance-degree matrices
> ➤ **edge distance index** ≡ *edge distance degree* → edge distance matrix

■ **edge distance matrix** ($^{\mathbf{E}}\mathbf{D}$)

Usually derived from the → *H-depleted molecular graph* $\mathcal{G}$, the edge distance matrix is the edge analogue of the vertex → *distance matrix*, and summarizes in matrix form the topological distance information among all the pairs of bonds [Bonchev, 1983; Estrada and Gutman, 1996]. It is simply the distance matrix of the → *line graph* of $\mathcal{G}$.

The **topological edge distance** $[^{\mathbf{E}}\mathbf{D}]_{ij}$ between the edges e_i and e_j is defined as the length of the shortest → *path* between them, that is, number of vertices in the shortest path connecting edges e_i and e_j not counting the terminal vertices of the path.

As each ith edge is characterized by two vertices $v_{i(1)}$ and $v_{i(2)}$ incident to the edge, the topological edge distance between the edges e_i and e_j can also be obtained from the minimum → *topological distance d* between two pairs of vertices as

$$[^{\mathbf{E}}\mathbf{D}]_{ij} = \min\{d_{i(1),j(1)}, d_{i(1),j(2)}, d_{i(2),j(1)}, d_{i(2),j(2)}\} + 1$$

For acyclic graphs, the topological edge distance can be calculated by the following formula:

$$[^{\mathbf{E}}\mathbf{D}]_{ij} = \frac{1}{4} \cdot \left(d_{i(1),j(1)} + d_{i(1),j(2)} + d_{i(2),j(1)} + d_{i(2),j(2)}\right)$$

The off-diagonal entries $[^{\mathbf{E}}\mathbf{D}]_{ij}$ of the edge distance matrix are equal to 1 if edges e_i and e_j are adjacent (i.e., the edges e_i and e_j are connected and $[^{\mathbf{E}}\mathbf{D}]_{ij} = [\mathbf{E}]_{ij} = 1$, where $[\mathbf{E}]_{ij}$ denotes the elements of the → *edge adjacency matrix* $\mathbf{E}$), otherwise they are more than 1. The diagonal elements are equal to zero. The edge distance matrix, usually derived from a → *H-depleted molecular graph*, is square symmetric with dimension $B \times B$, where B is the number of bonds.

The maximum value entry in the ith row is called **bond eccentricity** $^{\mathrm{b}}\eta_i$ (or **edge eccentricity**):

$$^{\mathrm{b}}\eta_i = \max_j\left([^{\mathbf{E}}\mathbf{D}]_{ij}\right)$$

From the eccentricity definition, a graph $\mathcal{G}$ can be immediately characterized by two molecular descriptors known as **topological radius from edge eccentricity** $^{\mathrm{E}}R$ and **topological diameter from edge eccentricity** $^{\mathrm{E}}D$. The radius is defined as the minimum bond eccentricity and the diameter as the maximum bond eccentricity, according to:

$$^{\mathrm{E}}R = \min_i\left(^{\mathrm{b}}\eta_i\right) \quad \text{and} \quad ^{\mathrm{E}}D = \max_i\left(^{\mathrm{b}}\eta_i\right)$$

From the edge distance matrix several → *topological information indices* are calculated. Moreover, the atomic and molecular descriptors already defined for the vertex distance matrix are analogously defined for the edge distance matrix.

From the frequencies of the matrix row entries, the **edge distance code** is defined as the ordered sequence of the occurrence of increasing edge distance values for the ith considered edge,

$$\{{}^{1}f_i, {}^{2}f_i, {}^{3}f_i, \ldots, {}^{b}\eta_i f_i\}$$

where ${}^{1}f_i, {}^{2}f_i, {}^{3}f_i, \ldots$, called **edge distance counts**, indicate the frequencies of the edge distances equal to 1, 2, 3, ..., respectively, from edge e_i to any other edge and ${}^{b}\eta_i$ is the ith bond eccentricity.

The **edge distance degree** (or **edge distance index**, **edge distance sum**) is the row sum ${}^{E}\sigma_i$ obtained by summing the ith row entries of the edge distance matrix:

$$ {}^{E}\sigma_i \equiv VS_i({}^{E}\mathbf{D}) = \sum_{j=1}^{B} [{}^{E}\mathbf{D}]_{ij} = \sum_{k=1}^{{}^{b}\eta_i} {}^{k}f_i \cdot k $$

where VS_i is the $\rightarrow$ *row sum operator* and ${}^{k}f_i$ is the edge distance count of kth-order, which runs over the different edge distance values.

The sum of the edge distance degrees, that is, the sum of all matrix elements, is called **total edge distance** D_E and defined as

$$ D_E = \sum_{i=1}^{B} \sum_{j=1}^{B} [{}^{E}\mathbf{D}]_{ij} = \sum_{i=1}^{B} {}^{E}\sigma_i $$

where B is the number of graph edges.

A $\rightarrow$ *Wiener-type index*, called **edge Wiener index** and denoted as ${}^{E}W$, can be obtained from the edge distance matrix ${}^{E}\mathbf{D}$ as

$$ {}^{E}W \equiv Wi({}^{E}\mathbf{D}) \equiv \frac{1}{2} \cdot \sum_{i=1}^{B} \sum_{j=1}^{B} [{}^{E}\mathbf{D}]_{ij} $$

where Wi is the $\rightarrow$ *Wiener operator* and B the number of graph edges [Gutman and Estrada, 1996]. It was demonstrated that ${}^{E}W$ differs from the standard $\rightarrow$ *Wiener index* only by a constant and, consequently, reflects the same molecular structure features, that is,

$$ {}^{E}W = W - \frac{A \cdot (A-1)}{2} $$

where A is the number of vertices in G, that is, the number of heavy atoms in the molecule considered.

More interesting is the $\rightarrow$ *edge-type Schultz index* derived from both the edge distance matrix and the edge adjacency matrix.

Bond multiplicity is taken into account by augmenting the edge distance matrix with a supplementary column and row for each multiedge, therefore obtaining an **edge distance matrix for multigraphs** [Bonchev, 1983]. All the local vertex invariants and molecular descriptors defined above can also be calculated on this matrix.

Moreover, if the chemical bond distances are calculated in real molecules, a **geometric edge distance matrix**, denoted as ${}^{E}\mathbf{G}$, can be obtained from which analogous geometric descriptors are derived. Each entry of this matrix is calculated from the interatomic distances between the end point vertices of the pair of edges considered:

$$ [{}^{E}\mathbf{G}]_{ij} = \frac{1}{4} \cdot \left(r_{i(1),j(1)} + r_{i(1),j(2)} + r_{i(2),j(1)} + r_{i(2),j(2)} \right) $$

where r is the Euclidean distance; $i(1)$ and $i(2)$ are the vertices incident to the ith edge, while $j(1)$ and $j(2)$ are the vertices incident to the jth edge.

The **reciprocal edge distance matrix** (or **edge Harary matrix**), denoted as $^\mathbf{E}\mathbf{D}^{-1}$, is a square symmetric $B \times B$ matrix whose off-diagonal entries are the reciprocal distances between edges considered:

$$[^\mathbf{E}\mathbf{D}^{-1}]_{ij} = \begin{cases} 1/[^\mathbf{E}\mathbf{D}]_{ij} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

The diagonal entries are zero by definition. This matrix is the $\rightarrow$ *reciprocal distance matrix* of the line graph of the molecular graph G [Ivanciuc, Ivanciuc *et al.*, 1997].

📖 [Estrada, Rodriguez *et al.*, 1997; Estrada and Rodriguez, 1997]

➢ **edge distance matrix for multigraphs** $\rightarrow$ edge distance matrix
➢ **edge distance sum** $\equiv$ *edge distance degree* $\rightarrow$ edge distance matrix
➢ **edge eccentricity** $\equiv$ *bond eccentricity* $\rightarrow$ edge distance matrix
➢ **edge-Gutman index** $\rightarrow$ Schultz molecular topological index
➢ **edge-Harary index** $\rightarrow$ weighted matrices ($\odot$ weighted adjacency matrices)
➢ **edge Harary matrix** $\equiv$ *reciprocal edge distance matrix* $\rightarrow$ edge distance matrix
➢ **edge-Hosoya matrix** $\rightarrow$ Hosoya Z matrix
➢ **edge layer matrix** $\rightarrow$ layer matrices
➢ **edge matrices** $\rightarrow$ matrices of molecules
➢ **edge orbital information content** $\rightarrow$ orbital information indices
➢ **edge radial centric information index** $\rightarrow$ centric indices
➢ **edges** $\rightarrow$ graph
➢ **edge-Schultz index** $\rightarrow$ Schultz molecular topological index
➢ **edge-Szeged matrices** $\rightarrow$ Szeged matrices
➢ **edge type count** $\rightarrow$ edge adjacency matrix
➢ **edge-valence-weighted Schultz distance matrix** $\rightarrow$ weighted matrices ($\odot$ weighted distance matrices)
➢ **edge-vertex incidence matrix** $\rightarrow$ incidence matrices
➢ **edge-vertex-valence-weighted Schultz distance matrix** $\rightarrow$ weighted matrices ($\odot$ weighted distance matrices)
➢ **edge-vertex-weighted Schultz distance matrix** $\rightarrow$ weighted matrices ($\odot$ weighted distance matrices)
➢ **edge-weighted Schultz distance matrix** $\rightarrow$ weighted matrices ($\odot$ weighted distance matrices)
➢ **edge-weighted detour matrix** $\equiv$ *weighted detour matrix* $\rightarrow$ detour matrix
➢ **edge-weighted Harary index** $\rightarrow$ weighted matrices ($\odot$ weighted adjacency matrices)
➢ **edge-weighted Harary matrix** $\rightarrow$ weighted matrices ($\odot$ weighted adjacency matrices)
➢ **edge Wiener index** $\rightarrow$ edge distance matrix
➢ **edge-Wiener matrix** $\rightarrow$ Wiener matrix
➢ **edge-XI matrix** $\rightarrow$ weighted matrices ($\odot$ weighted adjacency matrices)
➢ **edge-Zagreb matrix** $\rightarrow$ weighted matrices ($\odot$ weighted adjacency matrices)
➢ **edge-χ matrix** $\equiv$ χ *matrix* $\rightarrow$ weighted matrices ($\odot$ weighted adjacency matrices)
➢ **EEVA descriptors** $\rightarrow$ EVA descriptors

➢ **Effective Dose** → biological activity indices (⊙ pharmacological indices)
➢ **effective resonance constant** → electronic substituent constants (⊙ resonance electronic constants)
➢ **effective solute hydrogen-bond acidity** → Linear Solvation Energy Relationships (⊙ hydrogen-bond parameters)
➢ **effective solute hydrogen-bond basicity** → Linear Solvation Energy Relationships (⊙ hydrogen-bond parameters)
➢ **EFVCI** ≡ *External Factor Variable Connectivity Index* → variable descriptors
➢ **eigenvalue-based descriptors** ≡ *spectral indices*
➢ **EigenVAlue descriptors** ≡ *EVA descriptors*
➢ **eigenvalues** → algebraic operators (⊙ characteristic polynomial)
➢ **eigenvalues of the adjacency matrix** → spectral indices
➢ **eigenvalues of the distance matrix** → spectral indices
➢ **eigenvectors** → algebraic operators (⊙ characteristic polynomial)
➢ **eigenvector centrality** → center of a graph
➢ **EIM** ≡ *Electronic Indices Methodology* → quantum-chemical descriptors (⊙ EIM descriptors)
➢ **EIM descriptors** → quantum-chemical descriptors
➢ **electrical conductance matrix** ≡ *conductance matrix* → resistance matrix
➢ **electric dipole moment** ≡ *dipole moment* → electric polarization descriptors
➢ **electric permittivity** → physico-chemical properties (⊙ dielectric constant)

■ **electric polarization descriptors**
Electric polarization, dipole moments, and other related physical quantities, such as multipole moments and polarizabilities, constitute a group of both local and molecular descriptors, which can be defined either in terms of classical physics or in terms of quantum mechanics. They encode information about the charge distribution in molecules [Böttcher, van Belle *et al.*, 1973]. They are particularly important in modeling solvation properties of compounds that depend on solute/solvent interactions and in effect frequently used to represent the → *dipolarity/polarizability term* in → *Linear Solvation Energy Relationships*. Moreover, they can be used to model the polar interactions that contribute to determine → *lipophilicity* of compounds.

The **dipole moment μ** (or **electric dipole moment**) is a vectorial quantity that encodes displacement with respect to the center of gravity of positive and negative charges in a molecule, defined as

$$\boldsymbol{\mu} = \sum_i q_i \cdot \mathbf{r}_i$$

where q_i are point charges located at positions $\mathbf{r}_i$. The SI unit for dipole moments is the coulomb meter, but they are often expressed in debye.

The elements of the vector **μ** are called **dipole moment components**:

$$\mu_x = \sum_i q_i \cdot x_i \qquad \mu_y = \sum_i q_i \cdot y_i \qquad \mu_z = \sum_i q_i \cdot z_i$$

where x, y, z are the coordinates of the charges. Molecules with zero dipole moments are called *nonpolar*, others *polar*; moreover, dipole moments equal to zero indicate molecules with a center of symmetry.

In analogy to the definition of electric dipole moment, electric multipole moments are also defined. In particular, the **quadrupole moment Q** and the **octupole moment U** are defined as

$$\mathbf{Q} = \frac{1}{2!} \cdot \sum_{i=1}^{A} q_i \cdot \mathbf{r}_i \cdot \mathbf{r}_i \qquad \mathbf{U} = \frac{1}{3!} \cdot \sum_{i=1}^{A} q_i \cdot \mathbf{r}_i \cdot \mathbf{r}_i \cdot \mathbf{r}_i$$

where $\mathbf{Q}$ is a tensor of second degree (a 3×3 symmetric matrix) and $\mathbf{U}$ a tensor of third degree (a $3 \times 3 \times 3$ symmetric matrix).

For example, when the charge distribution is spherically symmetrical, all the diagonal terms of the quadrupole moment are equal to zero. Therefore, the trace of the electric quadrupole moment is a measure of molecular charge distribution deviation from sphericity.

When a molecule is embedded in a uniform electric field $\mathbf{E_0}$ *in vacuum*, an **induced dipole moment** $\boldsymbol{\mu}_{IND}$ arises, defined by the relationship:

$$\boldsymbol{\mu}_{IND} = \alpha \cdot \mathbf{E}_0$$

where the scalar constant of proportionality is called **polarizability** α (or **static polarizability**). This scalar polarizability may be regarded as the sum of the **electronic polarizability** α_E and the **atom polarizability** α_A. A polarizable molecule shows an induced dipole moment different from zero.

For substituent groups, the **excess electron polarizability** $\Delta\alpha_E$ was also defined [Dearden, Bradburne *et al.*, 1991] as the difference between the calculated electron polarizability for straight chain alkyl groups by using a model based on the $\rightarrow$ *McGowan characteristic volume* V_X and the effective electron polarizability of the substituent:

$$\Delta\alpha_E = \alpha_E - [0.135 \cdot V_X + 0.052]$$

In general, the scalar polarizability α is not sufficient to describe the induced polarization; therefore, a **polarizability tensor** $\vec{\boldsymbol{\alpha}}$ is used to better encode induced polarization and represents **molecular polarizability**. In such a general case, the induced dipole moment needs not to have the same direction as the applied field, but the direction will depend on the position of the molecule relative to the polarizing field [Miller and Savchik, 1979; Miller, 1990c].

Each molecule, polar or nonpolar, is polarizable, that is, its electrons can be shifted under an electric field $\mathbf{E}$ so that **polarization** P is induced in the molecule. The polarization is proportional to the electric field strength $\mathbf{E}$ and the simplest relationships between P and $\mathbf{E}$ are

$$P = \chi^e \cdot \mathbf{E} = \frac{\varepsilon - 1}{4\pi} \cdot \mathbf{E}$$

where the scalar proportionality constant χ^e is the $\rightarrow$ *dielectric susceptibility*. In the second equation, ε is the $\rightarrow$ *dielectric constant*.

Polarization can be factorized in two main contributions: **induced polarization** P_α, due to translation effects, and **dipole polarization** P_μ, due to orientation of permanent dipoles. Moreover, induced polarization can be viewed as being due to the contribution of **electronic polarization** P_E *and atomic polarization* P_A:

$$P = P_\alpha + P_\mu = P_E + P_A + P_\mu$$

Other important quantities related to polarization and dipole moments are listed below.

- **molar polarization**

The dipole moment induced per unit of volume V is called molar polarization P_M and is defined by the Clausius–Mossotti equation as

$$P_M = \frac{\varepsilon - 1}{\varepsilon + 2} \cdot \frac{MW}{\rho} = \frac{4\pi}{3} \cdot N_A \cdot \alpha = \frac{n_D^2 - 1}{n_D^2 + 2} \cdot \frac{MW}{\rho} = MR$$

where MW is the molecular weight, ρ the density, and ε the dielectric constant; in the second equation, N_A is the Avogadro number, α the scalar polarizability, and MR the → *molar refractivity*. For high frequency fields the relationship $\varepsilon = n^2$ holds, where n is the → *refractive index*; the subscript D indicates the value of the refractive index corresponding to the sodium D-line, as it is usually used.

- **atom polarizability**

This is the polarization effect at atomic level, where dipoles $\boldsymbol{\mu}_{IND,i}$ are induced on each atom as

$$\boldsymbol{\mu}_{IND,i} = \alpha_i \cdot \mathbf{E}_i$$

where $\mathbf{E}_i$ is the electric field at the ith atom and α_i the corresponding polarizability, assumed to be isotropic. Atom polarizabilities are linearly correlated with their → *hardness* [Politzer, 1987]. Atomic contributions to polarizability (Table E4) were estimated by several authors [Kang and Jhon, 1982; Miller, 1990b; No, Cho *et al.*, 1993] and are used to calculate the mean polarizability of a molecule by summing the atomic contributions.

Table E4 Atomic polarizability values: (1) Kang–Jhon atomic hybrid polarizabilities α_A(ahp); (2) Miller–Savchik average atomic polarizabilities α^*_A(ahc); (3) No–Cho–Jhon–Sheraga atomic polarizabilities $\alpha^*_{ij,0}$; and linear charge coefficient α_{ij}.

Atom/hybrid	1 α_A	2 α^*_A	3 $\alpha_{ij,0^*}$	4 α_{ij}
H bonded to Xsp_3	0.386	0.392	0.396	0.219
H bonded to Xsp_2	0.386	0.392	0.298	0.404
C sp_3	1.064	1.116	1.031	0.590
C sp_2 arom.	1.382	1.369	1.450	0.763
C sp_2 carbonyl			1.253	0.862
C sp_2 ethylene			1.516	0.568
C sp	1.279	1.294		
N sp_3	1.094	1.077	0.966	0.437
N sp_2 amide			0.821	0.422
N sp_2 pyrrole	1.090	0.851	0.871	0.424
N sp_2 pyridine	1.030	0.910	0.656	0.436
N sp	0.852	0.972		
O sp_3	0.664	0.780	0.623	0.281
O sp_2 carbonyl	0.460	0.739	0.720	0.347
O sp_2 aromatic	0.422	0.586	0.720	0.347
S sp_3	3.000[a]	3.056	2.688	1.319

(*Continued*)

Table E4 (*Continued*)

Atom/hybrid	1 α_A	2 α_A^*	3 $\alpha_{ij,0^*}$	4 α_{ij}
S sp$_2$ thione	3.729[a]	3.661		
S sp$_2$ aromatic	2.700[a]	2.223		
P sp$_3$	1.538[a]	1.647		
F	0.296[a]	0.527	0.226	0.144
Cl	2.315[a]	2.357	2.180	1.089
Br	3.013[a]	3.541	3.114	1.402
I	5.415[a]	5.573	5.166	2.573

Data from Miller [Miller, 1990b].

• **mean polarizability**

The molecular polarizability is a tensor when the molecule is not perfectly spherical. The mean polarizability α of a molecule is calculated by the relation

$$\alpha = \frac{\alpha_{xx} + \alpha_{yy} + \alpha_{zz}}{3}$$

where α_{xx}, α_{yy}, and α_{zz} are the polarizability along each principal component axis of the molecule, obtained by diagonalization of the polarizability tensor [Cartier and Rivail, 1987]. When, for practical purposes, the effect due to anisotropy of polarization is small, the mean polarizability is calculated simply as

$$\alpha = \frac{tr(\vec{\alpha})}{3}$$

where $tr(\vec{\alpha})$ is the trace of the nondiagonalized polarizability tensor $\vec{\alpha}$.

Molecular polarizability can also be approximated by simply summing atomic polarizabilities over all the molecule atoms. Moreover, molecular polarizability α expressed as polarizability volume is called **polarizability volume** and is defined as

$$\alpha' = \frac{\alpha}{4\pi\varepsilon_0}$$

where ε_0 is the $\rightarrow$ *dielectric constant in vacuum*.

• **atom–atom polarizability**

Index of chemical reactivity, denoted as π_{ab}, is among the $\rightarrow$ *quantum-chemical descriptors* and calculated from the perturbation theory as

$$\pi_{ab} \equiv AAP_{ab} = 4 \cdot \sum_{i=1}^{N_{OCC}} \sum_{j=1}^{N_{OCC}} \sum_{\mu} \sum_{\nu} \frac{c_{i\mu,a} \cdot c_{j\mu,a} \cdot c_{i\nu,b} \cdot c_{j\nu,b}}{\varepsilon_i - \varepsilon_j}$$

where i and j run over the molecular orbitals, ε_i and ε_j denote their corresponding energies, μ and ν run over the atomic orbitals of the atoms a and b, and c denotes the coefficients of the linear combination of atomic orbitals ϕ defining each molecular orbital ψ.

The **self-atom polarizability** π_{aa} is analogously defined as

$$\pi_{aa} \equiv SAP_{aa} = 4 \cdot \sum_{i=1}^{N_{OCC}} \sum_{j=1}^{N_{OCC}} \sum_{\mu} \sum_{\nu} \frac{c_{i\mu,a}^2 \cdot c_{j\nu,a}^2}{\varepsilon_i - \varepsilon_j}$$

An index for **total self-atom polarizability** is obtained as

$$^T\pi = \sum_{a=1}^{A} \pi_{aa}$$

where π_{aa} is the self-atom polarizability of the ath atom and A is the number of atoms in the molecule.

• **anisotropy of the polarizability** (β^2)
A measure of the deviation of the molecular polarizability from a spherical shape, defined as

$$\beta^2 = \frac{\left(\alpha_{xx}-\alpha_{yy}\right)^2 + \left(\alpha_{yy}-\alpha_{zz}\right)^2 + \left(\alpha_{zz}-\alpha_{xx}\right)^2}{2}$$

where α_{xx}, α_{yy}, and α_{zz} are the polarizabilities along each principal component axis of the molecule, obtained by diagonalization of the polarizability tensor $\vec{\vec{\alpha}}$.

Some empirical **polarity/polarizability descriptors,** which were proposed to measure the ability of the compound to influence a neighboring charge or dipole by virtue of dielectric interactions, are discussed below.

• **electrostatic factor**
A molecular descriptor proposed for solvent classification and defined as

$$EF = \varepsilon \cdot \mu$$

where ε is the dielectric constant and μ the magnitude of the dipole moment of the solvent.

In general, values between 0 and 2 indicate hydrocarbon solvents, between 3 and 20 electron-donor solvents, between 20 and 50 hydroxylic solvents, and values greater than 50, dipolar aprotic solvents.

• **Polarizability Effect Index** (PEI)
The polarizability effect index is based on the stabilizing energy E_X caused by the polarizability effect for a substituent X interacting with a point charge q [Cao and Li, 1998]. For alkyl and aliphatic alcohol substituents, the stabilizing energy E_X is defined as:

$$E_X = K \cdot \sum_i \left[N_i \cdot \frac{1 + \cos\theta}{q - \cos\theta} - \frac{2 \cdot \cos\theta \cdot (1 - (\cos\theta)^{N_i})}{(1 - \cos\theta)^2} \right]^{-2} = K \cdot \text{PEI}(X)$$

where $K = -2.16 q^2/(2\varepsilon r_{CC}^4)$, the constant -2.16 being the average polarizability α value of the four basic alkyl units (CH_3, CH_2-, $CH<$, and $-C<$), ε the effective dielectric constant and r_{CC} the carbon-carbon bond length; the summation goes over all the basic units of the substituent X. N_i is the number of heavy atoms between the ith alkyl unit and the probe with charge q; for each atom located at unit distance from the probe, N_i represents the topological distance from the probe. θ is the supplementary angle of bond angle CCC (i.e. $\theta = 180° - 109.5° = 70.5°$ for sp^3 hybridization). In the last term of the expression of E_X, PEI(X) is the relative order of

Table E5 Values of the ΔPEI contributions at different
topological distance from the point charge center.

N_i	ΔPEI	N_i	ΔPEI	N_i	ΔPEI	N_i	ΔPEI
1	1.000000	6	0.009052	11	0.002375	16	0.001073
2	0.140526	7	0.006388	12	0.001972	17	0.000945
3	0.048132	8	0.004748	13	0.001628	18	0.000838
4	0.025503	9	0.003666	14	0.001421	19	0.000749
5	0.013800	10	0.002916	15	0.001229	20	0.000673

polarizability effect of alkyl substituents, called polarizability effect index; this can be calculated as the following:

$$\text{PEI(X)} = \sum_i \Delta\text{PEI}_i(\text{X})$$

where $\Delta\text{PEI}_i(\text{X})$ is the contribution of the ith basic alkyl unit of the substituent. In Table E5, values of $\Delta\text{PEI}_i(\text{X})$ are reported for different values of N_i.

Taking each ith carbon atom of an alkane molecule as the beginning one in the action of the probe, a $\text{PEI}(\text{X}_i)$ value can be calculated in the same way as PEI(X) for the substituent X; then, the **Molecular Polarizability Effect Index** (MPEI) is derived as [Cao and Li, 1998]:

$$\text{MPEI} = \sum_{i=1}^{A} \text{PEI}(\text{X}_i)$$

where the summation is over all the carbon atoms in the molecule.

The **Geometric Mean Polarizability Effect Index** (GMPEI) was also proposed for alkanes as [Cao and Yuan, 2002]:

$$\text{GMPEI} = \left[\prod_{i=1}^{A} \text{PEI}(\text{X}_i) \right]^{1/A}$$

An analogous approach was proposed for alkene derivatives, where the **Geometric Mean Polarizability Effect Index of π bond** (GMPEIπ) [Cao and Yuan, 2002] is calculated taking only into account the PEI contributions for the π bond.

The **Inner Molecular Polarizability Index** (IMPI) was defined as [Cao, Liu *et al.*, 1999]:

$$\text{IMPI} = \sum_{i=1}^{A} \text{PEI}_i$$

where A is the number of carbon atoms in the molecule and PEI_i the sum of the polarizability effect index of the alkyl groups bonded to the i-th atom. This last index differs from MPEI because PEI_i is the polarizability effect index of alkyl groups connected to the i-th carbon atom, whereas $\text{PEI}(\text{X}_i)$ is the polarizability effect index of the whole molecule seen as a substituent X with the i-th carbon atom as the beginning one.
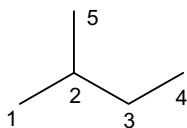
By comparing the IMPI value of an alkane with the IMPI value of the corresponding n-alkane, the quasi-length carbon chain $N_{\text{C(eff)}}$ was defined as:

$$N_{\text{C(eff)}} = \frac{\text{IMPI}_{n\text{-}alkane}}{\text{IMPI}_{isomer}} \cdot N_{\text{C}}$$

where N_{C} is the number of carbon atoms of the two isomers.

Example E3

Calculation of MPEI, GMPEI, and IMPI for 2-methylbutane.

C_1: $PEI(X_1) = 1.00000 + 0.14053 + 2 \times 0.04813 + 0.02350 = 1.2603$

C_2: $PEI(X_2) = 1.00000 + 3 \times 0.14053 + 0.04813 = 1.4697$

C_3: $PEI(X_3) = 1.00000 + 2 \times 0.14053 + 2 \times 0.04813 = 1.3773$

C_4: $PEI(X_4) = 1.00000 + 0.14053 + 0.04813 + 2 \times 0.02350 = 1.2357$

C_5: $PEI(X_5) = 1.00000 + 0.14053 + 2 \times 0.04813 + 0.02350 = 1.2603$

C_1: $PEI_1 = 1.00000 + 2 \times 0.14053 + 0.04813 + 0.02350 = 1.3292$

C_2: $PEI_2 = 3 \times 1.00000 + 0.14053 = 3.1405$

C_3: $PEI_3 = 2 \times 1.00000 + 2 \times 0.14053 = 2.2811$

C_4: $PEI_4 = 1.00000 + 0.14053 + 2 \times 0.04813 = 1.2368$

C_5: $PEI_5 = 1.00000 + 2 \times 0.14053 + 0.04813 = 1.3292$

$MPEI = 1.2603 + 1.4697 + 1.37773 + 1.2357 + 1.2603 = 6.6033$

$GMPEI = [1.2603 \times 1.4697 \times 1.37773 \times 1.2357 \times 1.2603]^{1/5} = 1.3177$

$IMPI = 1.3292 + 3.1405 + 2.2811 + 1.2368 + 1.3292 = 9.3167$

📖 [Cao, Yuan *et al.*, 2000, 2003; Liu, Liang *et al.*, 2006]

- **Kier–Hall solvent polarity index ($^1\chi_f^v$)**

It is a solvent polarity index defined as the first-order → *valence connectivity index* $^1\chi^v$ divided by the number N_f of discrete isolated functional groups to account for multiple interaction sites as [Kier and Hall, 1986]:

$$^1\chi_f^v = {}^1\chi^v / N_f$$

It was assumed that functional groups influencing solvent polarity are π-electron systems and lone pairs. Thus, for example, N_f (benzene) = 1, N_f (nitrobenzene) = 2 (one π-electron system and one lone pair), N_f (pyridine) = 2 (one π-electron system and one lone pair), N_f (nitro group) = 1. Halogen atoms are considered do not give a significant contribution to a molecule in enhancing its solvent polarity and thus for a halogen-containing molecule $^1\chi^v$ is used unmodified.

📖 [Sekusak and Sabljić, 1992]

- **local polarity index (Π)**

It is defined as the average deviation of the surface → *molecular electrostatic potential* calculated as [Brinck, Murray *et al.*, 1993; Murray, Brinck *et al.*, 1993]:

$$\Pi = \frac{1}{SA} \cdot \int\limits_0^{SA} |V(\mathbf{r}) - \bar{V}|\, dSA = \lim_{n \to \infty} \frac{1}{n} \cdot \sum_{i=1}^n |V(\mathbf{r}_i) - \bar{V}|$$

where $V(\mathbf{r})$ is the potential energy value at ith grid point on the molecular surface SA, $\bar{V}$ is average potential energy value on the surface for the considered molecule, and n is the number of grid points on the molecular surface.

Π ranges from zero for a neutral atom to 21.6 Kcal/mol for water. It is a measure of charge separation or local polarity; it has been shown to correlate with the $\rightarrow$ *dipolarity/polarizability term* Π^* as well as with the $\rightarrow$ *dielectric constant* ε. The local polarity index is among the descriptors used in the $\rightarrow$ *GIPF approach* and was used to calculate log P by the $\rightarrow$ *Politzer hydrophobic model*.

• **Q polarity index**
This is a topological polarity index derived by the electrotopological $\rightarrow$ *intrinsic state I* of the atoms in the molecule and defined as [Kier and Hall, 1999b]:

$$Q = \frac{A^2 \cdot \sum\limits_{i=1}^A I_i^{ALK}}{\left(\sum\limits_{i=1}^A I_i\right)^2}$$

where A is the number of atoms, I_i^{ALK} is the intrinsic state of the ith atom in the skeleton structure of the molecule in which each nonhydrogen atom is replaced with an sp^3 carbon atom on the corresponding isoconnective alkane as reference structure, and I_i is the intrinsic state of the ith atom in the actual compound. The basic idea is that Q value for the considered molecule lies between two extremes of minimal and maximal polarity: the minimal polarity is given by a molecule only constituted of sp^3 carbon atoms and the maximal polarity is approximated by the square of the number of atoms A in the molecule.

• **polar hydrogen factor** (Q_H)
This is an index of the molecular polarity due to C−H bonds restricted to halogenated hydrocarbons [Di Paolo, Kier *et al.*, 1979]. It is calculated as the sum of the contributions to the polarity of all the C−H bonds in a molecule. For each C−H bond, three different contributions are considered due to halogens linked to the same carbon atom of the C−H bond, halogens in α-position and halogens in β-position with respect to the considered C−H bond:

$$Q_H = \sum_b \left[\sum_C k_C + \sum_\alpha k_\alpha + \sum_\beta k_\beta \right]$$

where the external summation runs over all the C−H bonds and the three internal summations run over all halogens directly attached to the carbon atom of C−H bond, in α- and β-positions, respectively. When no halogen is attached to a C−H bond, its contribution to Q_H is taken as zero. The only exception is made for a methylene group CH_2 flanked by two halogen-substituted methyl groups: in this case C−H bonds are considered in the calculation.

The constant values k (Table E6) are defined according to the Swain–Lupton $\rightarrow$ *field-inductive constant* $\mathcal{F}$, that is, they represent the relative field effect of an atom or a group.

Table E6 *k* parameters for the polar hydrogen factor Q_H.

Halogen	k Contribution	Halogen	k Contribution
F	0.43	α-Cl	0.10
Cl	0.41	α-Br	0.09
Br	0.44	β-F	0.05
I	0.40	β-Cl	0.05
α-F	0.13	β-Br	0.05

For examples, Q_H values for $CHCl_3$, CH_2Cl_2, CF_3-CHFCl, and CF_3-CH_2-CF_2Cl are $3 \times 0.41 = 1.23$, $4 \times 0.41 = 1.64$, $0.43 + 0.41 + 3 \times 0.13 = 1.23$, $2 \times (5 \times 0.13 + 0.10) = 1.50$, respectively.

📖 [Hannay and Smyth, 1946; McClellan, 1963; Buckingham, 1967; Böttcher, van Belle *et al.*, 1973; Exner, 1975; Lien, Liao *et al.*, 1979; Lien, Guo *et al.*, 1982; Li, Guo *et al.*, 1984; Topsom, 1987c; Lewis, 1989; Miller, 1990c; Beck, Glen *et al.*, 1996; Stuer-Lauridsen and Pedersen, 1997; Beck, Horn *et al.*, 1998; Norinder, Sjöberg *et al.*, 1998; Norinder, Österberg *et al.*, 1999; Hansch, Steinmetz *et al.*, 2003]

➤ **electromeric effect** → electronic substituent constants
➤ **electron affinity** → quantum-chemical descriptors
➤ **electron charge density weight** → connectivity indices (⊙ charge-weighted vertex connectivity indices)
➤ **Electron-Conformational Approach** → Electronic-Topological method
➤ **Electron-Conformational Matrix of Congruity** → Electronic-Topological method
➤ **Electron-Conformational method** → Electronic-Topological method
➤ **Electron-Conformational Submatrix of Activity** → Electronic-Topological method
➤ **electronic delocalization entropy** → MARCH-INSIDE descriptors
➤ **electron density** → quantum-chemical descriptors
➤ **electron donor-acceptor substituent constant** → electronic substituent constants

◼ **electronegativity**
Atom electronegativity is among the most important → *atomic properties*. The concept of atom electronegativity was recognized as a useful basic principle in chemistry more than 150 years ago [Pritchard and Skinner, 1955]. Originally defined by Pauling [Pauling, 1939], electronegativity is *"the power of an atom in a molecule to attract electrons to itself."*

The classical definition of atomic electronegativity is due to Mulliken [Mulliken, 1934, 1955a, 1955b] – **Mulliken electronegativity** –:

$$\chi^{MU} = \frac{IP + EA}{2}$$

that is, the arithmetic mean of the → *ionization potential* IP and the → *electronic affinity* EA of the atom. In this definition, the use of ionization potentials and electron affinities of valence states was proposed.

The Mulliken scale of electronegativity (in volts) can be converted in the Pauling scale χ^{PA} (*Pauling units*) by the empirical relation:

$$\chi^{PA} = 0.303 \cdot \chi^{MU}$$

Electronegativity scales, other than Pauling and Mulliken scales, are the Allred–Rochow scale χ^{AR} [Allred and Rochow, 1958, 1961] based on estimated effective nuclear potentials and covalent radii; the Gordy scale χ^{G} [Gordy, 1946, 1951] based on the number of electrons in the valence shell of the atom and the covalent radius; the Sanderson scales χ^{SA} [Sanderson, 1952, 1954, 1955, 1971] based on covalent radii; the Hinze–Jaffé scale χ^{HJ} [Hinze and Jaffé, 1962, 1963b, 1963a; Hinze, Whitehead *et al.*, 1963] based on orbital energies and effective charges; Zhang scale χ^{Z} based on ionization energies and covalent radii [Zhang, 1982b, 1982a]. Atom electronegativity can also be estimated by functions of the → *vertex degree*, as proposed by Kier–Hall (→ *Kier–Hall electronegativity*) and by Roy–Ghosh in the framework of → *ETA indices* [Roy and Ghosh, 2003]. An extended review of electronegativity scales was published by Luo and Benson [Luo and Benson, 1990] and some of them are listed in Table E7.

Table E7 Electronegativity values from different sources (Pauling units).

	H	B	C	N	O	F	Si	P	S	Cl	Br	I
Pauling 1	2.1	2.0	2.5	3.0	3.5	4.0	1.8	2.1	2.5	3.0	2.8	2.5
Pauling 2	2.20	2.04	2.55	3.04	3.44	3.98	1.90	2.19	2.58	3.16	2.76	2.66
Mulliken	2.28	2.01	2.63	2.33	3.17	3.91	2.44	1.81	2.41	3.00	2.76	2.56
Allred–Rochow	2.20	2.01	2.50	3.07	3.50	4.10	1.74	2.06	2.44	2.83	2.74	2.21
Sanderson 1	2.31	1.88	2.47	2.93	3.46	3.92	1.74	2.16	2.66	3.28	2.96	2.50
Sanderson 2	2.592	2.275	2.746	3.194	3.654	4.000	2.138	2.515	2.957	3.475	3.219	2.778
Mullay	2.08	1.85	2.47	3.41	3.15	4.00	1.91	1.99	2.49	3.07	2.81	2.47
Gordy	2.17		2.52	2.98	3.45	3.95			2.58	3.50	2.75	2.50
Wells	2.28		2.30	3.35	3.70	3.95			2.80	3.03	2.80	2.47
Boyd–Markus	1.94	1.95	2.53	3.23	3.53	4.00	1.81	2.34	2.65	3.14	2.78	2.48
Inamoto–Masuda	2.00		2.21	2.71	3.02	3.05	1.72	1.93	2.15	2.37	2.32	2.15
Diudea	1.680	1.501	1.831	2.240	2.680	3.024	1.424	1.646	2.026	2.512	2.279	1.879
Zhang	2.271	1.966	2.536	3.062	3.642	4.188	1.769	2.131	2.479	2.835	2.529	2.142

(1) Pauling 1 [Pauling, 1939]; (2) Pauling 2 [Allred and Rochow, 1961]; (3) Mulliken [Mulliken, 1934, 1935a]; (4) Allred–Rochow [Allred and Rochow,1958]; (5) Sanderson 1 [Sanderson, 1952, 1955]; (6) Sanderson 2 [Sanderson, 1988]; (7) Mullay [Mullay, 1984]; (8) Gordy [Gordy, 1946]; (9) Wells [Wells, 1968a]; (10) Boyd–Markus [Boyd and Markus, 1981]; (11) Inamoto–Masuda [Inamoto and Masuda, 1982]; (12) Diudea [Diudea, Kacso *et al.*, 1996]; (13) Zhang [Zhang, 1982a].

The concept of electronegativity has become increasingly general (perhaps, even ambiguous) during its revision in the different quantum-chemistry frameworks, ranging from atomic electronegativity to orbital and functional group electronegativity up to molecular electronegativity. Electronegativity is a property of the state of the system; electrons tend to flow from a region of low electronegativity to a region of high electronegativity. With the formation of a

molecule, electronegativities of the constituent atoms or fragments equalize, all becoming equal to the electronegativity of the final state of the molecule (**Sanderson's electronegativity equalization principle** [Sanderson, 1951, 1971; Zefirov, Kirpichenok *et al.*, 1987]).

The concept of electronegativity equalization led to the calculation of dipole moments [Malone, 1933; Ferreira, 1963a], bond dissociation energies [Ferreira, 1963b], atomic charges [Stoklosa, 1973], and force constants [Gordy, 1946; Polansky and Derflinger, 1963]. Moreover, based on electronegativity, several → *bond ionicity indices* were proposed. To avoid chemically unacceptable results due to the total equalization of the electronegativity in a molecule [Gasteiger and Marsili, 1980], a partial equalization principle was also taken into account as a result of changes in orbital overlap [Pritchard, 1963]. In particular, a **partial equalization of orbital electronegativity (PEOE)** for the calculation of partial atomic charges was proposed using a topological iterative approach [Gasteiger and Marsili, 1980; Marsili and Gasteiger, 1980]; a **modified partial equalization of orbital electronegativity (M-PEOE)** was also successively proposed [No, Grant *et al.*, 1990a, 1990b]. Moreover, equalization based on → *density functional theory* was also proposed [Itskowitz and Berkowitz, 1997].

Group electronegativity was proposed for molecular substituents, functional groups, and fragments, where the center atom is the atom connected to the body of the molecule.

Group electronegativity is defined as the electronegativity of the central atom of the substituent and is affected by the neighbors of the central atom. Several methods have been proposed to estimate group electronegativities [Clifford, 1959; McDaniel and Yingst, 1964; Huheey, 1965; Huheey, 1966; Inamoto and Masuda, 1982; Mullay, 1984, 1985; Bratsch, 1985; Xie, Sun *et al.*, 1995]. For example, **Sanderson group electronegativity** ESG is calculated as the geometric mean of the electronegativities of atoms belonging to the considered group [Sanderson, 1983]:

$$\text{ESG}_i = \left(\chi_1^{\text{SA}} \cdot \chi_2^{\text{SA}} \cdot \ \cdots \ \cdot \chi_m^{\text{SA}} \right)^{1/m}$$

where χ^{SA} is the Sanderson electronegativity and m the number of atoms of the ith molecular group. Valence group carbon-related electronegativities EC were derived from Sanderson's ESG values by taking into account the covalent radii (as mean bond lengths) relative to the sp^3 carbon atom [Diudea, Kacso *et al.*, 1996].

Another approach to the calculation of group electronegativities χ_{G} is based on a stepwise addition method [Zhou, Nie *et al.*, 2007]:

$$\chi_{\text{G}}(i) = \frac{1}{n_{i1}} \cdot \sum_{j=1}^{n_{i1}} \chi_{j1} \quad \text{where} \quad \chi_{j1} = \frac{1}{n_{j2}} \cdot \sum_{k=1}^{n_{j2}} \chi_{k2} \cdots \chi_{m(L-1)} = \frac{1}{n_{mL}} \cdot \sum_{q=1}^{n_{mL}} \chi_{qL}$$

where i represents the group focused atom, n_{i1} is the number of atoms directly bonded to the ith atom, χ_{j1} is the Pauling electronegativity of the jth atom at distance one from the ith atom, n_{j2} is the number of atoms directly bonded to the jth atom, which are at distance two from the ith atom, χ_{k2} is the Pauling electronegativity of the kth atom at distance two from the ith atom and bonded to the jth atom, and so on, until the last level corresponding to a distance equal to L from the ith atom (Figure E1).
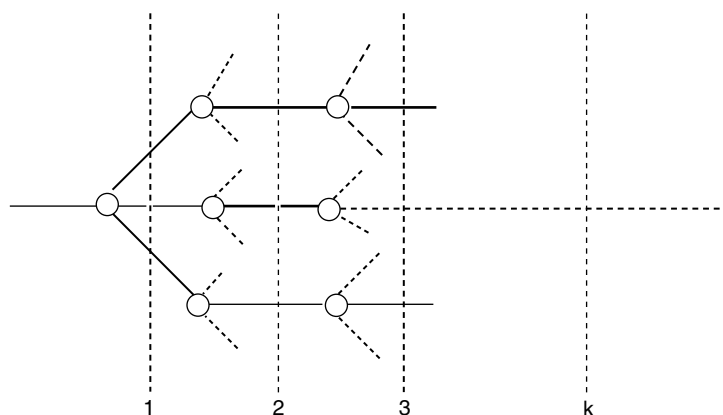
Figure E1 Levels of a group structure.

The equilibrium electronegativity is then defined as

$$\chi_i^{Eq} = \frac{\chi_i + \sum \chi_G(j)}{1+k}$$

where the summation goes over the electronegativity of the groups directly bonded to the *i*th atom and *k* is the number of these contributions.

Electronegativity values, derived from the Pauling electronegativity scale, are shown in Table E8 for some chemical groups.

Table E8 Electronegativity values for some chemical groups from Zhou, Nie *et al.* [Zhou, Nie *et al.*, 2007].

Groups	χ_G	Groups	χ_G	Groups	χ_G	Groups	χ_G
$-CH_3$	2.2875	$-CH_2CH_3$	2.3094	$-CH=CH_2$	2.3556	$-C\equiv CH$	2.4625
$-CN$	2.7950	$-NCO$	3.0100	$-OCN$	3.1175	$-C_6H_5$	2.4333
$-CHO$	2.7300	$-COOH$	2.9467	$-SiH_3$	2.1250	$-NH_2$	2.4800
$-NO$	3.2400	$-NO_2$	3.3067	$-OPh$	2.9367	$-OCH_3$	2.8638
$-OH$	2.8200	$-CCl_3$	3.0075	$-COCH_3$	2.7592	$-COPh$	2.8078

Based on equilibrium electronegativity, some molecular descriptors were proposed as, for example, the → *Nt index*.

Using group electronegativities as local invariants → *electronegativity-based connectivity indices* were proposed as molecular descriptors. Atomic electronegativities are also used in the definition of → *MARCH-INSIDE descriptors*.

📖 [Boyd and Edgecombe, 1988; Diudea and Silaghi-Dumitrescu, 1989b; Cherkasov, Galkin *et al.*, 1999, 2000; Cherkasov, 2003; Gilson, Gilson *et al.*, 2003; Agrawal, Gupta *et al.*, 2005; Leyssens, Geerlings *et al.*, 2005]

➤ **electronegativity-based inductive constant** → **electronic substituent constants** (⊙ inductive electronic constants)

■ **electronegativity-based connectivity indices**

These are molecular descriptors defined by analogy with the → *Randić connectivity index* and calculated on hydrogen-included molecular graphs where heavy vertices are weighted by valence group electronegativities [Diudea, Kacso *et al.*, 1996].

The first proposed electronegativity-based connectivity index is the **DSI index** [Diudea and Silaghi-Dumitrescu, 1989a] based on the → *Sanderson group electronegativity* ESG_i used as the → *local vertex invariant*. It is defined as

$$DSI = \sum_{b=1}^{B} (EVG_i \cdot EVG_j)_b^{-1/2}$$

where the summation goes over all edges in the graph.

EVG_i is the valence group electronegativity of the ith vertex calculated from the Sanderson group electronegativity (geometric mean of electronegativities of the atoms belonging to the considered group G_i), accounting for atom valence as

$$EVG_i = (ESG_i)^{1/(1+\delta_i)} = \left[\left(\chi_i^{SA} \cdot \left(\chi_h^{SA} \right)^{h_i} \right)^{1/(1+h_i)} \right]^{1/(1+\delta_i)}$$

where χ_i^{SA} and χ_h^{SA} are the Sanderson electronegativities of the ith heavy atom and the hydrogen atom, respectively, δ_i is the → *vertex degree* of the ith atom, and h_i is the number of hydrogen atoms belonging to the group G_i of the ith vertex and calculated as

$$h_i = 8 - L_i - \delta_i$$

where L is the principal quantum number. When $\delta_i > 8 - L_i$, then $h_i = 0$ by definition; moreover, if multiple bonds are present, the vertex degree should be replaced by the sum of the conventional bond orders (i.e., → *bond vertex degree* δ_i^b).

An extension of DSI index to paths of higher order was also proposed as

$$^{m}DSI = \sum_{^{m}p_{ij}} (EVG_i \cdot EVG_k \cdot \ldots \cdot EVG_j)^{-1/2}$$

where the sum is over all the paths of length m in the graph and the product over the valence group electronegativities of all the vertices involved in each path.

Other two electronegativity-based connectivity indices are the **ECP index** and **ECN index** based on the valence group carbon-related electronegativities EC_i proposed by Diudea [Diudea, Kacso *et al.*, 1996]:

$$ECP = \sum_{i=1}^{A} ecp_i = \sum_{i=1}^{A-1} \sum_{j=i+1}^{A} a_{ij} \cdot (EC_i \cdot EC_j)^{1/2}$$

$$ECN = \sum_{i=1}^{A} ecn_i = \sum_{i=1}^{A-1} \sum_{j=i+1}^{A} a_{ij} \cdot (EC_i \cdot EC_j)^{-1/2}$$

where ecp_i and ecn_i are local vertex invariants calculated by summing the products of the electronegativity of the considered ith vertex by the electronegativities of its bonded atoms, a_{ij} denotes the elements of the → *adjacency matrix* equal to 1 for adjacent vertices and otherwise zero, and A is the number of heavy atoms in the molecule.

> ➢ **electronegativity ETA measure** → ETA indices
> ➢ **electronegativity scales** → electronegativity
> ➢ **electronegativity-weighted adjacency matrix** → weighted matrices (⊙ weighted adjacency matrices)
> ➢ **electronegativity-weighted edge connectivity index** → edge adjacency matrix
> ➢ **electronegativity-weighted walk degrees** → walk counts
> ➢ **electronic charge index** ≡ *total absolute atomic charge* → charge descriptors
> ➢ **electronic chemical potential** → quantum-chemical descriptors

■ **electronic descriptors**

These are local or global molecular descriptors related to the electronic distribution in the molecule; they are fundamental to many chemical reactions, physico-chemical properties, and ligand–macromolecule interactions. The theory of electronic density is based on a quantum-mechanical approach; however, → *electronegativity* and charges, which are not physical observables, are also important quantities for the definition of several electronic descriptors.

A lot of → *quantum-chemical descriptors* are derived from the charge distribution in a molecule or the → *electron density* of specified atoms or molecular regions, and from conformational energy values such as the → *Joshi electronic descriptors*. Several → *charge descriptors* and → *electric polarization descriptors* are calculated from atomic charge estimations.

Electronic information is combined with shape and steric information to characterize molecules in → *charged partial surface area descriptors*. Other approaches, different from those closely related to quantum-chemistry, refer to electronic distribution in molecules, such as → *electronic substituent constants*, → *electrotopological state indices*, → *topological charge indices*. → *Reactivity indices* and → *delocalization degree indices* are also related to electronic properties of molecules.

A simple example of electronic descriptors is the **lone-pair electron index**, denoted by LEI, calculated from the → *H-depleted molecular graphs* as [Cheng and Yuan, 2006]:

$$\text{LEI} = \frac{1}{A} \cdot \sum_{i=1}^{n_{\text{het}}} \left[\text{LE}_i \cdot \sum_{j=1}^{A} \frac{1}{d_{ij}^2} \right] \quad j \neq i$$

where the first summation runs over the heteroatoms and the second one over all the heavy atoms; A is the number of heavy atoms, d_{ij}^2 is the square → *topological distance* between atoms v_i and v_j, and LE_i is the **lone-pair electrostatic interaction** of the ith heteroatom, defined as

$$\text{LE}_i = \frac{\sqrt{L_i} \cdot \left(Z_i^v - Z_i^b \right)}{\chi_i^{\text{PA}}}$$

where L is the principal quantum number, χ^{PA} the Pauling → *electronegativity*, Z^v the number of valence electrons, and Z^b the number of bonding electrons.

The values of the lone-pair electrostatic interaction are provided in Table E9.

Table E9 Lone-pair electrostatic interaction values (LE) of atoms.

Atoms	LE	Atoms	LE
C	0	Cl	3.2887
N	0.9304	Br	4.0541
O	1.6444	I	5.0438
F	2.1320		

➤ **Electronic EigenValue descriptors** ≡ *EEVA descriptors* → EVA descriptors
➤ **Electronic Indices Methodology** → quantum-chemical descriptors (⊙ EIM descriptors)
➤ **electronic polarizability** → electric polarization descriptors
➤ **electronic polarization** → electric polarization descriptors

■ **electronic substituent constants** (≡ *Hammett substituent constants,* σ *electronic constants*) Derived from the → *Hammett equation,* σ electronic constants are calculated for different molecular substituents from the rate or equilibrium constant of specific reactions, with respect to a reference compound [Topsom, 1976, 1987b; Charton, 1981; Taft and Topsom, 1987].

These substituent descriptors are usually used in → *Hansch analysis* as the molecular electronic properties in the case of a monosubstituted series of compounds, while they are summed up over all the molecule substituents to obtain global molecular descriptors in the general case of polysubstituted compounds.

The possible modes of action of substituents in modifying the electron distribution of the parent molecule can be distinguished in two main effects: a **polar effect** (or **field-inductive effect** or **localized effect**) and a **resonance effect** (or **delocalized effect**) (Figure E2). The term polar effect is used to characterize the influence of unconjugated, sterically remote substituents on equilibrium or rate processes. The polar substituent effect is transmitted through-bonds (**inductive effect** or **static inductive effect**) involving a polarization either of the σ-bond network (**σ-inductive effect**) or of the π-bond network (**π-inductive effect**). It can also be transmitted through solvent space according to classical laws of electrostatics (**field effect**) by the formation of bond dipole moments. Transmission efficiency of the localized effects is empirically measured by a **transmission coefficient** ε ($0 \leq \varepsilon \leq 1$).
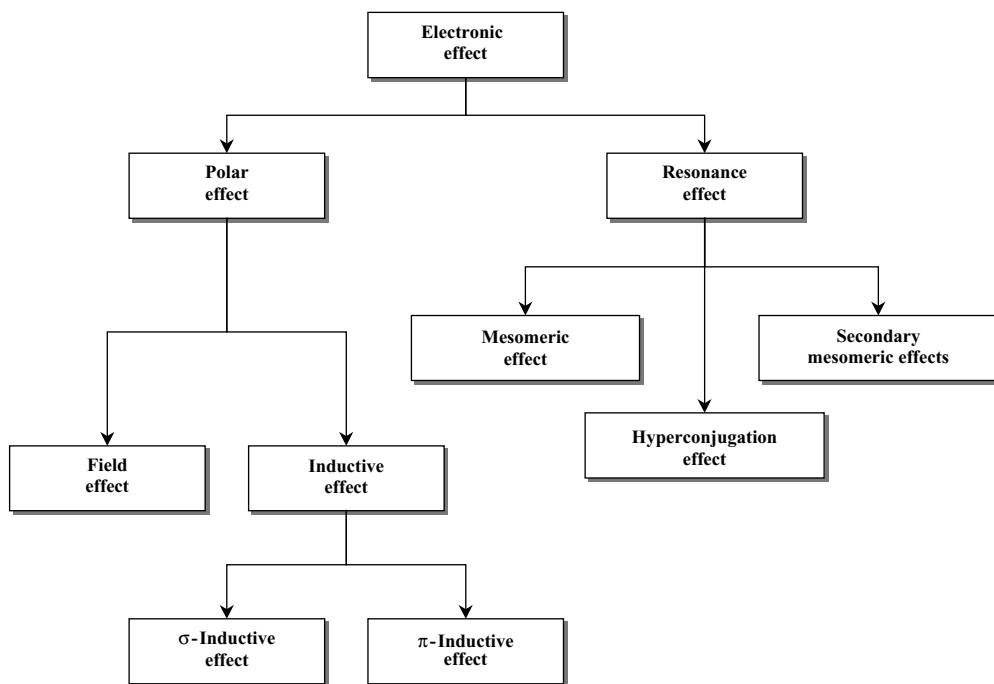

Figure E2 Scheme of the relationships among the electronic effects.

Note that the separability of the two contributions of the polar effect is difficult to attain; in effect, attempts to separate the polar effect contributions have been unsuccessful so they are usually considered together. However, from a theoretical point of view, field effects have been studied using the Kirkwood–Westheimer [Kirkwood and Westheimer, 1938; Westheimer and Kirkwood, 1938] and Tanford models [Tanford, 1957].

Resonance effect is an energy stabilization caused by delocalization of electrons in the bond network of the molecule and can be attributed to a **mesomeric effect**, that is, the delocalization of π electrons on the π orbital network, a **hyperconjugation effect**, that is, a delocalization of σ electrons in a π orbital aligned with the σ bond, and **secondary mesomeric effects**, such as repulsion of the π electrons by nonbonded electrons on a substituent or solvent, or by time-dependent effects due to polarizabilities (for the last, the term **electromeric effect** is sometimes used).

If the substituent is bonded to an sp^3 carbon atom not involved in the π molecular orbitals formation, its electrical effect is only local, assuming that σ electron delocalization is negligible. Substituents bonded to sp or sp^2 carbon atoms can exert both localized and delocalized effects.

Taking into account the basic contributions defined above, the overall electronic effect σ of the substituent can be represented by the following equation:

$$\sigma = \ell \cdot \sigma_I + d \cdot \sigma_R$$

where σ_I and σ_R are the polar and resonance contributions, respectively; ℓ and d weight their importance in determining the overall effect. Equivalent expressions with the same meaning are represented by different symbols by other authors, such as

$$\sigma = f \cdot \mathcal{F} + r \cdot \mathcal{R} \quad \text{or} \quad \sigma = \lambda \cdot \sigma_D + \delta \cdot \sigma_L$$

From these coefficients, the percentage of the resonance effect $D\%$ can be obtained as

$$D\% = \frac{d}{d + \ell} \cdot 100$$

Several electronic substituent constants were defined so as to represent both global and particular electronic effects. The σ values obtained unambiguously from experimentally accessible data or from the many possible reaction series are called *primary values* and the corresponding set *primary standard*. The σ values derived from the primary values, by rescaling with modified ρ constants or correlation equations, are called *secondary values* and the corresponding set *secondary standard*.

The most popular electronic substituent descriptors are listed below; Table E10 collects the main information concerning all the electronic substituents.

Table E10 Summary of the electronic substituent constants.

ID	Symbol	Reaction	ρ	D%	References
1	σ_p	K; 4-XC$_6$H$_4$COOH	1.00	53	[McDaniel and Brown, 1958]; Lewis and Johnson, 1959]
2	σ_m	K; 3-XC$_6$H$_4$COOH	1.00	22	[McDaniel and Brown, 1958]; [Lewis and Johnson, 1959]
3	σ_p^+	k; 4-XC$_6$H$_4$C(CH$_3$)$_2$Cl	−4.54	66	[Brown and Okamoto, 1958]

(Continued)

Table E10 *(Continued)*

ID	Symbol	Reaction	ρ	*D*%	References
4	σ_m^+	k; 3-$XC_6H_4C(CH_3)_2Cl$	−4.54	33	[Brown and Okamoto, 1958]
5	σ_p^-	K; p-XC_6H_4OH	2.23	56	[Lewis and Johnson, 1959; Hine, 1962; Cohen and Jones, 1963]
6	σ_m^0	K and k; unbiased values	1.00	23	[Taft, 1960]
7	$\bar{\sigma}_p$	K and k; effective values	1.00	42	[Taft, 1960]
8	σ^0	k; HO^- + $XC_6H_4CH_2COOEt$	0.98	37	[Yukawa, Tsuno *et al.*, 1966]
9	σ_p^n	K and k; "normal" values	1.00	47	[van Bekkum]
10	σ^*	k; XCH_2COOY	2.48	6	[Taft, 1952, 1953b, 1956]; [Taft and Lewis, 1958]
11	σ^*	k; $XCOOY$	2.48	37	[Taft, 1952, 1953b, 1956]; [Taft and Lewis, 1958]
12	σ^*	k; $X_2CHCOOY$	2.48	0	[Taft, 1952, 1953b, 1956]; [Taft and Lewis, 1958]
13	σ^*	k; $X_2(CH_2)_2COOY$	2.48	15	[Taft, 1952, 1953b, 1956]; [Taft and Lewis, 1958]
14	σ^*	k; $X_2(CH_2)_3COOY$	2.48	33	[Taft, 1952, 1953b, 1956]; [Taft and Lewis, 1958]
15	σ^*	k; 2-XC_6H_4COOY	1.00	53	[Taft, 1952, 1953b, 1956]; [Taft and Lewis, 1958]
16	σ'	K; 4-$XC_8H_{12}COOH$	1.464	3	[Roberts and Moreland, 1953]
17	σ'	K; 4-$XC_8H_{12}COOH$	1.65	1	[Holtz and Stock, 1964]; [Baker, Parish *et al.*, 1967]
18	σ_I	$\sigma_I = 0.45\,\sigma^*$ k; XCH_2COOY	6.23	0	[Taft, 1952, 1953b, 1956]; [Taft and Lewis, 1958]
19	σ_R	$\sigma_R = \sigma - \sigma_I$	1.00	92	[Taft, Ehrenson *et al.*, 1959]; [Taft and Lewis, 1959]; [Taft Price *et al.*, 1963a]
20	σ_R^0	$\sigma_R^0 = \sigma^0 - \sigma_I$	1.00	84	[Taft, Ehrenson *et al.*, 1959]; [Taft and Lewis, 1959]; [Taft, Price *et al.*, 1963a]
21	σ_I^q	k; $XC_7H_{12}N$	1.00	—	[Grob and Schlageter, 1976]
22	σ''	K; 4-$XC_6H_{10}COOH$	1.00	0	[Siegel and Komarmy, 1960]
23	σ_j^{DG}	K; j-$XC_{10}H_8COOH$	1.46	29	[Dewar and Grisdale, 1962b]
24	σ_I	$\sigma_I = b \cdot pK_a^R + a$	—	—	Charton, 1964]
25	σ_m^I	ionization potential, $XC_6H_4CH_2^{\bullet}$	1.00	26	[Harrison, Kebarle *et al.*, 1961]
26	σ_p^I	ionization potential, $XC_6H_4CH_2^{\bullet}$	1.00	65	[Harrison, Kebarle *et al.*, 1961]
27	σ^Q	$\sigma^Q = (f - 34.826)/1.024$ ^{35}Cl quadrupole resonance	1.00	26	[Bray and Barnes, 1957]
29	σ_p^F	^{19}F nmr chemical shift, XC_6H_4F	1.00	65	[Taft, 1960]
30	σ_p^C	^{13}C nmr chemical shift, XC_6H_5	1.00	68	[Maciel and Natterstad, 1965]

(Continued)

Table E10 *(Continued)*

ID	Symbol	Reaction	ρ	D%	References
31	ι (iota)	$\iota = 1.755 - \delta^C/54.9$	—	—	[Inamoto and Masuda, 1977]; [Inamoto, Masuda *et al.*, 1978]
		^{13}C nmr shift, XC_6H_4Y			
32	σ_χ	Theoretical calculations	—	—	[Marriott and Topsom, 1982]; [Marriott, Reynolds *et al.*, 1984]
33	I	*K*; acids	1.00	5	[Branch and Calvin, 1941]
34	σ_R^0	$\sigma_R^0 = 0.0079\,A^{1/2} - 0.027$	1.00	96	[Brownlee, Katritzky *et al.*, 1965]; [Brownlee, Katritzky *et al.*, 1966]
		Infrared spectroscopy			
35	σ_R^0	$\sigma_R^0 = (\delta_p^F - \delta_m^F)/2.97$	—	—	[Taft, Ehrenson *et al.*, 1959]
		^{19}F nmr chemical shift, XC_6H_4F			
36	R	*K*; 4-X-pyridinium ions	1.00	—	[Taft and Grob, 1974]
37	*F*	$f(\sigma_m, \sigma_p)$	1.00	22	[Dewar and Grisdale, 1962b]
38	*M*	$f(\sigma_m, \sigma_p)$	1.00	93	[Dewar and Grisdale, 1962b]
39	*F′*	$f(\sigma_m, \sigma_p)$	1.00	27	[Dewar and Grisdale, 1962b]
40	*M′*	$f(\sigma_m, \sigma_p)$	1.00	93	[Dewar and Grisdale, 1962b]
41	σ_p^+	$f(\sigma_m, \sigma_p)$	1.00	70	[Swain and Lupton Jr., 1968]
42	$\mathcal{F}$	$f(\sigma_m, \sigma_p)$	1.00	0	[Swain and Lupton Jr., 1968]
43	$\mathcal{R}$	$f(\sigma_m, \sigma_p)$	1.00	100	[Swain and Lupton Jr., 1968]
44	$\Delta\sigma$	$\sigma_p - \sigma_m$	1.00	92	[McDaniel and Brown, 1958]; [Lewis and Johnson, 1959]
45	C_T	X-cianoethylenes	—	—	[Hetnarski and O'Brien, 1975]
46	σ^ϕ	Dialkylphosphinic acids	—	—	[Mastryukova and Kabachnik, 1971]
47	σ_a	Heterocyclic rings	—	—	[Otsuji, Kubo *et al.*, 1960]

The symbols *K* and *k* in the reaction column represent equilibrium $(1/\rho)\,\log\,(K_X/K_H)$ values and kinetic $(1/\rho)\,\log\,(k_X/k_H)$ values, respectively; subscripts X and H refer to the X- and H-substituted compounds, respectively. For all the series the unsubstituted compounds have σ = 0 except for series 15–17, 27, 29, and 30, whose values are: (15) σ = 7.760; (16) σ = 7.760; (17) σ = 34.622; (27) σ = 0.490; (29) σ = −0.100; (30) σ = −0.115. The values of the sensitivity to resonance effect (D%) are taken from Swain–Lupton [Swain and Lupton Jr.,

1968]. σ_j electronic constants of Dewar–Grisdale (37–40) represent σ values determined from *j*-$XC_{10}H_8COOH$ substituted compounds (1-naphtoic acids) with *j* = 3, 4, 5, 6, 7; the corresponding percentages of resonance (D%) are 29, 57, 38, 43, and 48. σ^Q values (27) were estimated from the ^{35}Cl quadrupole resonance frequency *f* in *ortho*-substituted chlorobenzenes; it is related to the Taft σ^* polar constant. σ_R^0 values (34) were estimated from the integrated intensity *A* of the ν_{16} band in IR spectra.

• overall electronic constants σ_m and σ_p

These are the original Hammett substituent constants [Hammett, 1937, 1970] measuring the overall electronic effect of the *meta*- and *para*-substituents of benzene derivatives having the functional group in the side chain. They were originally calculated from the variation of the acid

dissociation constant K_a of substituted benzoic acids (m-, p-XC_6H_4COOH) in water at $25\,°C$, with respect to the unsubstituted compound (i.e., benzoic acid):

$$\sigma_{m,p} = \frac{1}{\rho} \cdot \log\left(\frac{K^X}{K^0}\right) = \frac{1}{\rho} \cdot (pK_a^0 - pK_a^X)$$

where the reaction constant ρ is arbitrarily assumed equal to 1. K^X and K^0 denote the acid dissociation constants of the X-substituted and parent compound, respectively. The subscripts m and p are used to identify *meta*- and *para*-substituents, respectively. Other values of $\sigma_{m,p}$ were also obtained from the hydrolysis of benzoic esters and other reaction series, based on both equilibrium and rate constants (K and k, respectively).

σ values measure the substituent total electronic effect with respect to hydrogen and are, in principle, independent of the nature of the reaction; however, for large and/or charged substituents the σ values are estimated with less precision.

- **unbiased constants σ^0**

The unbiased constants σ^0 were defined by Taft to avoid overestimating the resonance effect caused by direct resonance interactions between substituent and reaction site [Taft, 1960], that is, they are a measure only of the interaction between the substituent and the molecular skeleton as felt at the reaction site.

These σ^0 constants were evaluated by the dissociation of *meta*-substituted phenylacetic acids and esters and were defined for a selected group of m-XC_6H_4- substituents which exhibit a precise linear free energy relationship. Their values are independent of whether the process is rate or equilibrium, the solvent and reaction conditions.

They represent inductive constants for the *meta*-substituted phenyl groups (m-XC_6H_4-) relative to the unsubstituted phenyl group (C_6H_5-), because substituent X in the *meta* position is not directly conjugated with the side chain reaction center Y, thus specific -C_6H_4Y resonance effects do not contribute to the overall electronic effect. However, σ^0 values contain contributions from resonance interactions within m-XC_6H_4- groups. The same does not hold for the corresponding *para*-substituted phenyl substituents (p-XC_6H_4-). Therefore, by using the linear free energy relationships for selected *meta*-substituted groups to determine the ρ reaction constant, the effective $\bar{\sigma}$ values were obtained for all p-XC_6H_4-:

$$\bar{\sigma} = 1/\rho \cdot \log(k/k_0)$$

The difference $\sigma - \sigma^0$ can be considered as a measure of the resonance effect between the aromatic group and the reaction center Y for the dissociation of *meta*-substituted benzoic acids in water. Moreover, deviations from the relationship $\log(k/k_0) = \sigma^0\rho$ provide a useful measure of the specific polar and resonance effects dependent upon both solvent conditions and reaction type.

- **inductive electronic constants**

These are electronic substituent constants representing the polar effect exerted by the substituent on the active site of a molecule. Different reaction series and data derived from various statistical procedures led to several proposals for inductive constants; only the most popular ones are considered in the following discussion.

The **Roberts–Moreland inductive constant** σ' is based on the dissociation constants K of 4-substituted bicyclo[2,2,2]octane-1-carboxylic acids in 50% ethanol by volume at $25\,°C$

defined as

$$\sigma' = \frac{1}{1.464} \cdot (\log K_X - \log K_H)$$

where $\rho = 1.464$ is assumed in the Hammett equation [Roberts and Moreland, 1953].

As the X substituent in bicyclo[2,2,2]octane-1-carboxylic acids is bonded to a sp^3 carbon atom, it exerts only an inductive effect. Moreover, the chosen reference compound is free from conformational effects and no steric effect is observed, as the substituent and the active site are not in close proximity to each other.

The **Holtz–Stock inductive constant** σ' was calculated by the dissociation constants K of 4-substituted bicyclo[2,2,2]octane-1-carboxylic acids in 50% ethanol by weight at 25 °C using $\rho = 1.65$ in the Hammett equation [Holtz and Stock, 1964].

Taft σ^* constant (or **σ^* electronic constant**, **Taft polar constant**) was proposed by Taft [Taft, 1956] to measure the inductive effect in the aliphatic series:

$$\sigma^* = \frac{1}{2.48} \cdot \left[\log\left(\frac{k_X}{k_{Me}}\right)_B - \log\left(\frac{k_X}{k_{Me}}\right)_A \right] = \frac{1}{2.48} \cdot \left[\log\left(\frac{k_X}{k_{Me}}\right)_B - E_S \right]$$

where A and B, respectively, stand for acid-catalyzed and base-catalyzed, k_X and k_{Me} denote the rate constants of acid- and base-catalyzed hydrolysis or esterification of substituted and unsubstituted esters, respectively, and E_S is the $\rightarrow$ *Taft steric constant*.

The factor 2.48 corresponds to the average of the available ρ values of alkaline hydrolysis from the Hammett equation and attempts to place σ^* on the same scale as the σ_m and σ_p electronic constants. Since base-catalyzed hydrolysis involves both inductive and steric effects and acid-catalyzed hydrolysis involves only the steric effect, removing the steric effect leaves only the inductive effect, assuming that in both reactions the steric effects are the same.

When defining the σ^* values as a measure of inductive effects, the choice in reactivity types is such that specific steric, resonance, and other effects are apparently constant and a linear free energy relationship of the Hammett type holds.

An **Additive Model of Inductive Effect** was proposed to estimate the σ^* inductive constant of substituents on the basis of the fundamental characteristics of the constituent atoms [Cherkasov, Galkin *et al.*, 1998; Cherkasov and Jonsson, 1998; Cherkasov, 2005]:

$$\sigma^* = \sum_{i=1}^{n} \frac{\sigma_i^A}{r_i^2} = 7.84 \times \sum_{i=1}^{n} \frac{\Delta\chi_i \cdot R_i^2}{r_i^2}$$

where the sum runs over all the atoms of the substituent, r_i is the distance of the *i*th atom of the substituent to the reaction center, and R_i is the covalent radius of the atom. σ^A is an empirical atomic parameter reflecting the ability of an atom to attract (or donate) electrons and its values were estimated by multivariate regression analysis on Taft σ^* constants for several substituents. The values were found to have good correlation with the difference in electronegativity $\Delta\chi$ between a given atom and the reaction center, reflecting the driving force of electron density displacement and, with the square of the covalent radius of the atom, reflecting the ability to delocalize the charge.

The **Taft–Lewis inductive constant** σ_I was proposed [Taft and Lewis, 1958] to measure the inductive effect in aliphatic series on a scale for direct comparison with aromatic σ values, derived from σ^* constant as

$$\sigma_I = 0.45 \times \sigma^* = \frac{1}{5.51} \times \left[\log\left(\frac{k_X}{k_H}\right)_B - \log\left(\frac{k_X}{k_H}\right)_A \right]$$

The derived ρ value of 5.51 was later modified by Taft into $\rho = 6.23$ [Taft, 1960]. In this way, σ_I values can also be considered as a measure of the inductive effect of substituents bonded to aromatic carbons.

These σ_I values of Taft and Lewis were used as a basis set by Charton [Charton, 1963, 1964] to obtain a large number of inductive constants. Acid dissociation constants of substituted acetic acids (XCH_2COOH) in water were correlated with σ_I constants of the basis set at temperatures from $5°$ to $50\,°C$ in terms of the equation:

$$\sigma_I = b \cdot pK_a^X + a$$

The regression coefficients a and b were estimated separately for each reaction series, and then additional σ_I values (**Charton inductive constants**) were estimated and a set of recommended values also suggested. In general, steric and resonance effects in the acetic acid system can be considered negligible. Moreover, unlike the Taft inductive constants, those defined by Charton require only one experiment for their determination.

The **Siegel–Kormany inductive constant** σ'' was calculated by the dissociation constants of 4-substituted cyclohexanecarboxylic acids in three different solvents (in water, water/ethanol 50% by weight and by volume) at $25\,°C$ using $\rho = 1$ in the Hammett equation [Siegel and Komarmy, 1960]. The expected relationship between this inductive constant and the inductive constant of Roberts–Moreland or Taft σ^* was confirmed.

Based on the acid dissociation of 4-substituted quinuclidines in water at $25°C$, the **Grob inductive constant** σ_I^q was proposed [Taft and Grob, 1974; Grob and Schlageter, 1976; Grob, 1985]:

$$\sigma_I^q = pK_a^X - pK_a^H$$

assuming $\rho = 1$ in the Hammett equation. The quinuclidine system is free of steric and conformational effects. Moreover, it is much more sensitive to electronic effects than the bicyclooctane system.

[19]F inductive constant σ_m^F was estimated from ^{19}F-NMR for *meta*-substituents of F-benzene in very dilute CCl_4 solution [Taft, 1960; Taft, Price *et al.*, 1963a] by the equation:

$$\sigma_m^F \equiv \sigma_I = 0.084 - \frac{\delta_m^F}{7.1}$$

where δ_m^F is the ^{19}F chemical shift of *meta*-substituted fluorobenzenes. The NMR chemical shifts measure the inductive perturbation of *meta*-substituents on the charge density of the fluorine neighbor. This relationship is based on the correlation between chemical shift and the Taft–Lewis inductive constant σ_I.

The **Inamoto–Masuda inductive constant** ι (*iota*) [Inamoto and Masuda, 1977] was empirically defined by modifying the electronegativity as defined by Gordy [Gordy, 1946]. The inductive

constant ι is based on the $\rightarrow$ *electronegativity* of the substituent atom directly bonded to the skeletal group and defined as

$$\iota = \frac{Z_{\text{eff}} + 1}{L_{\text{eff}}}$$

where Z_{eff} is the effective nuclear charge in the valence shell of the considered atom and L_{eff} is the effective principal quantum number (Slater rule). If the atom belongs to the second period group, the inductive constant is estimated by the equation:

$$\iota = 0.64 \cdot \chi_X + 0.53$$

where χ_X is the substituent charge obtained from the bond dipole moment [Inamoto, Masuda *et al.*, 1978].

The **electronegativity-based inductive constant** σ_χ was derived from atomic charge densities on the hydrogen atom of XH, XCH_2H, and XCH_2CH_2H derivatives [Marriott, Reynolds *et al.*, 1984] defined as

$$\sigma_\chi = 1 - q_H$$

where q_H is the atomic charge on the H atom and is calculated from $\rightarrow$ *computational chemistry*. The electronegativity of a substituent is primarily determined by the electronegativity of the nearest attached atom. Moreover, the electronegativity of the atom is affected by both changes in hybridization and the polarity of other atoms in the group. The values $1 - q_H$ being related to $\rightarrow$ *electronegativity scales*, such values can be considered a measure of the inductive effect of the substituent.

• **resonance electronic constants**

These are electronic substituent constants representing the resonance effect exerted by the substituent on the active site of a molecule. Several proposals of resonance constants were made on the basis of different reaction series or derived from statistical procedures; only the most popular are considered in the following.

The **Taft resonance constants** were calculated from the overall electronic effect in specific reaction series by subtracting the inductive contribution based on the Taft–Lewis σ_I values [Taft and Lewis, 1959; Taft, Ehrenson *et al.*, 1959; Taft, 1960; Ehrenson, Brownlee *et al.*, 1973] as

$$\sigma_R^0 = \sigma^0 - \sigma_I \qquad \bar{\sigma}_R = \bar{\sigma} - \sigma_I \qquad \sigma_R^+ = \sigma^+ - \sigma_I \qquad \sigma_R^- = \sigma^- - \sigma_I$$

In particular, σ_R^0 called **resonance polar effect** [Taft, 1956] is defined for any benzene derivative where there is no direct conjugation between substituent and reactive center; it can be considered constant for a particular solvent, therefore expressing resonance interactions between substituent and skeletal group. $\bar{\sigma}_R$ is usually referred to as the **effective resonance constant**; σ_R^+ and σ_R^- hold for electrophilic and nucleophilic reaction series, respectively.

Moreover, Taft also tried to propose a general σ_R-scale as

$$\sigma_R = \sigma - \sigma_I$$

However, the σ_R values for *para*-substituents show great variability with reaction type, and it was therefore inferred that a widely applicable and precise σ_R-scale could not be devised [Taft and Lewis, 1958, 1959].

Resonance constants σ_R^0 were also estimated from ^{19}F-NMR based on ^{19}F chemical shifts in *para*- and *meta*-substituted F-benzene in very dilute CCl_4 solution [Taft, Ehrenson *et al.*, 1959] by the equation:

$$\sigma_R^0 = \frac{1}{2.97} \cdot \left(\delta_p^F - \delta_m^F \right)$$

where δ^F is the ^{19}F chemical shift. The quantity $\delta_p^F - \delta_m^F$ can be considered a measure of the perturbation in electron density detected by the fluorine atom caused by the resonance interaction between the *para*-substituent and fluorobenzene system; δ_m^F is mainly related to the perturbation derived from bond polarizations, and the inductive contributions to δ_m^F and δ_p^F values are assumed to be equal.

Analogously, a significant correlation was found between δ_p^C (and $\delta_p^C - \delta_m^C$) and σ_R^0 or σ_R, δ^C being the chemical shift of the ^{13}C atom of the aromatic ring in the *meta*- or *para*-position relative to the substituent [Maciel and Natterstad, 1965].

Dual electronic constants σ^+ and σ^- were proposed to measure the "exaltation" of the resonance effect that appears when the substituent and active site bonded to a skeletal group give origin to a direct conjugation between them.

The **electrophilic substituent constant σ^+** measures the electronic effects for electron-releasing substituents (e.g., $-$OMe, $-$Me, $-$OH, $-$NH$_2$) and for a strong electron-acceptor active site, while the **nucleophilic substituent constant σ^-** measures the electronic effects for electron-acceptor substituents (e.g., $-$NO$_2$, $-$CN, $-$CO$_2$H) and strong electron-releasing active site.

Standard σ^+ values were estimated by Brown – Okamoto [Brown and Okamoto, 1958] from the rate constants k of the solvolysis of substituted *t*-cumyl chlorides (XC$_6$H$_4$C(CH$_3$)$_2$Cl) in 90% acetone–water at 25 °C, as

$$\sigma^+ = \frac{1}{-4.54} \cdot \log \frac{k}{k_0}$$

where the reaction constant $\rho = -4.54$ was estimated from a set of *meta*-substituents.

The σ^+ values correlate with ionization potentials obtained from substituted benzyl radicals [Harrison, Kebarle *et al.*, 1961].

Standard σ_p^- values were obtained from acid dissociation constants K_a of *para*-substituted phenols (*p*-XC$_6$H$_4$OH) in water or water/ethanol 50% at 25 °C [Cohen and Jones, 1963] and successively from those of *para*-substituted anilines.

When strong resonance interactions are less relevant, σ^+ and σ^- constants are equal to the normal σ values obtained from substituted benzoic acids.

The **Yukawa–Tsuno equation** (also referred to as **Linear Aromatic Substituent Reactivity relationship**, *LASR*) modifies the Hammett equation, taking into account the exaltation of the resonance effects of electron-releasing and electron-attracting substituents on the reaction center [Yukawa, Tsuno *et al.*, 1972a, 1972b].

For electron-releasing substituents, the Yukawa–Tsuno equation is

$$\log\left(\frac{k_X}{k_H}\right) = \rho \cdot \left[\sigma^0 + r \cdot (\sigma^+ - \sigma^0)\right] = \rho \cdot \left[\sigma^0 + r \cdot \Delta\bar{\sigma}^+\right]$$

where k_X and k_H are the respective rate constants of the X-substituted and unsubstituted compounds, ρ is the Hammett reaction constant, and r is the contribution of the enhanced resonance effect of the substituent. σ^0 is the "normal" (i.e., unbiased) substituent constant derived by Yukawa *et al.* from the rate constants of alkaline hydrolysis of ethyl phenyl-acetates ($HO^- + XC_6H_4CH_2COOEt$) in water at 25 °C.

$\Delta\sigma^+$ was proposed as the substituent constant measuring the exaltation of the resonance effect of a *para*-substituent on an electrophilic reaction, while the first term $\rho\sigma^0$ accounts primarily for the electronic effect of *meta*- and *para*-substituents whose σ^0 and σ^+ are equivalent.

Analogously, for electron-attracting substituents, the Yukawa–Tsuno equation is

$$\log\left(\frac{k_R}{k_0}\right) = \rho \cdot \left[\sigma^0 + r \cdot (\sigma^- - \sigma^0)\right] = \rho \cdot \left[\sigma^0 + r \cdot \Delta\bar{\sigma}^-\right]$$

Both these equations were originally proposed using σ values instead of σ^0 values [Yukawa and Tsuno, 1959]. If $r = 0$, the Yukawa–Tsuno equations reduce to the classical Hammett equation, while, if $r = 1$, it corresponds to the correlation with only σ^+ or σ^- constants.

A resonance constant R was calculated from two distinct reaction series as

$$R = \log\left(\frac{K}{K_0}\right)_I - \log\left(\frac{K}{K_0}\right)_{II}$$

where the subscript I represents the series of dissociation constants of 4-substituted pyridinium ions in water at 25 °C and II the series of dissociation constants of 4-substituted quinoclidinium ions in water at 25 °C [Taft and Grob, 1974].

This resonance constant is justified by the following correlation equations found separately for the two series:

$$\log\left(\frac{K}{K_0}\right)_I = 5.15 \cdot \sigma_I + 2.69 \cdot \sigma_R^+$$

$$\log\left(\frac{K}{K_0}\right)_{II} = 5.15 \cdot \sigma_I$$

Since the inductive effects are essentially the same in the two series, it follows that the difference in $\log(K_X/K_0)$ values for corresponding substituents in I and II reaction series gives a measure of the resonance effect for the dissociation of 4-substituted pyridinium ions.

- **field/resonance effect separation**

To give a uniform view of the different σ-scales, considering both field-inductive and resonance effect, a number of proposed approaches were aimed at separating the two main contributions within a unique theoretical framework.

According to the **Swain–Lupton approach** (SL), the σ electronic constant is defined as a linear combination of the two basic electronic contributions based on the equation [Swain and

Lupton Jr., 1968; Swain, Unger *et al.*, 1983; Reynolds and Topsom, 1984; Swain, 1984]:

$$\sigma^{SL} = f \cdot \mathcal{F} + r \cdot \mathcal{R}$$

where $\mathcal{F}$ is the **field-inductive constant** (or **Swain–Lupton field constant**) and $\mathcal{R}$ is the **resonance constant** (or **Swain–Lupton resonance constant**), f and r being weighting factors that depend on the system used to define the particular σ-scale (analogous to the coefficients ℓ and d defined above).

Swain and Lupton defined the field constant $\mathcal{F}$ assuming that the polar effect was a component in both Hammett electronic constants σ_m and σ_p:

$$\mathcal{F} = b_0 + b_1 \cdot \sigma_m + b_2 \cdot \sigma_p$$

where the coefficients were evaluated by least square regression using pK_a values of 4-substituted bicyclo[2,2,2]octane-1-carboxylic acids of Holtz–Stock. While Swain–Lupton assumed $\rho = 1$ so as to put the $\mathcal{F}$ values on the same scale as Hammett constants, Hansch used σ' as originally calculated by $\rho = 1.65$, thus obtaining the following equation [Hansch, Leo *et al.*, 1973]:

$$\mathcal{F} \equiv \sigma' = -0.009 + 1.369 \cdot \sigma_m - 0.373 \cdot \sigma_p$$

The resonance constant $\mathcal{R}$ was estimated by

$$\mathcal{R} = \sigma_p - 0.921 \cdot \mathcal{F}$$

assuming $r = 1$ and the coefficient f was evaluated assuming that $\mathcal{R} = 0$ for $N^+(CH_3)_3$. The main assumption is that the substituents in the *para*-position give the primary resonance effect.

Based on the same previous assumptions, but using a different ρ value ($\rho = 1.56$) and an extended data set of σ_I values, the following equation was used to define the field constant $\mathcal{F}$ and the corresponding resonance constant $\mathcal{R}$:

$$\mathcal{F} \equiv \sigma_I = 0.033 + 1.297 \cdot \sigma_m - 0.385 \cdot \sigma_p$$

Several sets of resonance constants were defined according to the different types of σ_p values. Resonance substituent constants $\mathcal{R}^+$, $\mathcal{R}^-$, and $\mathcal{R}^0$ are derived from the corresponding σ values as

$$\mathcal{R}^+ = \sigma_p^+ - f \cdot \mathcal{F} \quad \mathcal{R}^- = \sigma_p^- - f \cdot \mathcal{F} \quad \mathcal{R}^0 = \sigma_p^0 - f \cdot \mathcal{F}$$

using the appropriate field constant $\mathcal{F}$ values calculated by Swain–Lupton (the coefficient f is usually taken as equal to 1).

Also in the Taft–Lewis approach, the resonance R and polar effects I can be viewed as additive contributions to the overall electronic effect, defined as

$$\log\left(\frac{k}{k_0}\right) = I + R = \rho_I \cdot \sigma_I + \rho_R \cdot \sigma_R$$

where the inductive effect I is defined as in the Hammett equation.

This equation is applied separately to the effects of *meta*- and *para*-substituents, both depending on σ and ρ positions.

However, assuming

$$\rho_I = \rho_I^m = \rho_I^p = \rho_R^m = \rho_R^p = \rho_R$$

for *meta-* and *para-*substituents of benzene derivatives, the two following equations were proposed:

$$\log\left(\frac{K_p}{K_0}\right) = \rho_I \cdot (\sigma_I + \sigma_R)$$

$$\log\left(\frac{K_m}{K_0}\right) = \rho_I \cdot (\sigma_I + \alpha \cdot \sigma_R)$$

Moreover, the inductive constants are assumed equal in both cases while the resonance constant of the substituent in *meta-*position is considered a fraction of its resonance effect in the *para-*position. The coefficient α was originally proposed as equal to 1/3 for the dissociation of benzoic acids; other values were proposed to account for enhanced resonance effects.

By combining the two expressions, a general equation to measure the inductive effect was proposed as

$$\rho_I \cdot \sigma_I = \left(\frac{1}{1-\alpha}\right) \cdot \left[\log\frac{K_m}{K_0} - \alpha \cdot \log\frac{K_p}{K_0}\right]$$

According to the original **Dewar–Grisdale approach** (DG), also called **FM method** (Field and Mesomeric method) [Dewar and Grisdale, 1962a, 1962b], delocalized and localized long-range field effects are represented by the quantities M and F, respectively. Electronic substituent constants σ_{ij} of a given substituent X, bonded at position i of the skeletal group, acting on a reaction center bonded at position j, are defined by the equation:

$$\sigma_{ij}^{DG} = \frac{F}{r_{ij}} + M \cdot q_{ij}$$

where q_{ij} and r_{ij} represent the charge in position j and the distance (in units of benzene bond length) between i and j atoms. F and M values are calculated from σ_m and σ_p constants by substituting appropriate values in the equation. Once F and M values are known for each substituent, σ_{ij}^{DG} constants can be calculated for any structure at the i–j positions (e.g., *ortho*, *meta*, *para* positions in benzene derivatives, or other positions in naphthalene derivatives), using appropriate values of r_{ij} and q_{ij}. The term q_{ij} is taken as a measure of the transmission of mesomeric effects between position i and j of a conjugated system. In this approach, both substituent and reaction site are approximated by single point charges located at points i and j (Figure E3).
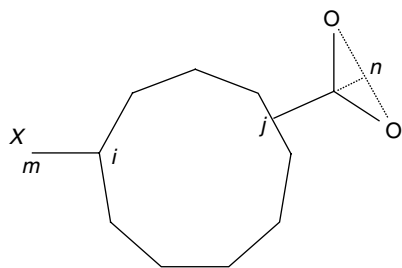

Figure E3 Geometrical scheme for the calculation of the Dewar–Grisdale electronic constants.

Substituting the charge term q_{ij} by the atom–atom polarizability term α_{ij} results in an analogous set of F' and M' values.

The **Dewar–Golden–Harris approach** (DGH), also called **FMMF method** (i.e., Field, Mesomeric, and Mesomeric–Field method) is a modification of the Dewar–Grisdale approach, where the substituent X is approximated by a finite dipole (represented by two point charges along the $i-X$ bond) and the reaction site at position j as a single point charge [Dewar, Golden *et al.*, 1971]. This approach is based on the following equation:

$$\sigma_{ij}^{DGH} = \frac{F}{r_{ij}} + M \cdot q_{ij} + M_F \cdot \sum_{k \neq j} \frac{q_{ik}}{r_{kj}}$$

where F and M, respectively, represent localized and delocalized effects of the substituent X, and q_{ik} is the charge of the carbon atom at different positions k in the skeletal moiety, and r is the distance (in units of benzene bond length) between atom j and any other atom in the skeletal group. The third substituent parameter M_F is the mesomeric–field constant, describing the ability of the substituent to polarize adjacent π systems; if there is no direct resonance interaction between the substituent and the reaction center, M_F should be proportional to M.

In the event of there being substituent effects on the dissociation of the carboxylic acids, the previous equation becomes

$$\sigma_{ij}^{DGH} = F \cdot R_{ij} + M \cdot q_{ij} + M_F \cdot \sum_{k \neq j} \frac{q_{ik}}{r_{kn}}$$

where the quantity R_{ij} is defined as

$$R_{ij} = \frac{1}{r_{in}} - \frac{0.9}{r_{mn}}$$

where m is a point charge (with charge $q = -0.9$) at distance 1.40 Å from the point charge i (with charge $q = 1$) along the $i-X$ bond; n is the point charge in the middle of the axis joining the two oxygens of the carbonyl group of the reaction site.

Further modifications of two previous approaches were proposed by Forsyth [Forsyth, 1973], calculating σ_{ij}^+ constants.

Other substituent electronic constants were defined for specific different reference compounds (e.g., heterocycles) and reactions.

Aryl electronic constants σ_a are electronic constants defined for substituents on an aromatic ring different from benzene, such as pyridine [Otsuji, Kubo *et al.*, 1960]. The concept may be generalized to various isocyclic and heterocyclic rings such as thiophene, furan, and so on, and various types of aryl electronic constants σ_a^0, σ_a^+, σ_a^- may be also defined. A particular set of this kind is given by σ_r **electronic constants** obtained by the protonation reaction of hydrocarbons, assuming as the reference compound α-naphtyl ($\sigma_r = 0$).

Phosphorus electronic constants σ^ϕ are electronic substituent constants derived from the dissociation constants of dialkylphosphinic acids for substituent groups directly bonded to a phosphorus atom [Mastryukova and Kabachnik, 1971]. Assuming the alkyl groups exert only an inductive effect, these electronic constants can be distinguished as σ_I^ϕ inductive constant and σ_R^ϕ resonance constant [Charton and Charton, 1978].

Radical electronic constants are substituent constants derived from free-radical reactions. The most popular are the E_R **radical parameter** defined on the basis of the radical abstraction of α-hydrogens of substituted cumenes [Yamamoto and Otsu, 1967] and the σ_{α}^{*} **radical substituent constants** defined by the benzylic α-hydrogen hyperfine coupling constants [Wayner and Arnold, 1984].

The **charge transfer constant** C_T (or **group charge transfer**) is an electronic substituent constant defined from the dissociation constant of a complex between tetracyanoethylene and a X-substituted parent compound [Hetnarski and O'Brien, 1975]:

$$C_T = \log K_X - \log K_H$$

This definition is analogous to that used for σ electronic substituent constants and accounts for the formation ability of a charge-transfer complex (CTC), such as the π-complex formation ability of aromatic systems.

The **electron donor–acceptor substituent constant** κ is an electronic constant proposed to measure the ability of a group to modify the stability of an electron donor–acceptor complex that is often referred to as charge transfer complex [Foster, Hyde *et al.*, 1978; Livingstone, Hyde *et al.*, 1979]. It is defined as

$$\kappa = \log K_X^{App} - \log K_H^{App}$$

where K^{App} is the apparent equilibrium constant for the formation of a complex between the electron-acceptor 1,3,5-trinitrobenzene and the X-substituted benzene in CCl_4 solution at 33.4 °C. Such equilibrium constants were determined by the NMR technique. Benzene was chosen as the reference compound.

📖 Additional references are provided in the thematic bibliography (see Introduction).

➢ **electronic-topological descriptors** → charge descriptors
➢ **Electronic-Topological Matrix of Conjunction** → Electronic-Topological method

■ **Electronic-Topological method** (≡ *ET method*)
This is a QSAR approach based on a matrix representation of chemical compounds involving geometrical and electronic features. This approach is mainly aimed at identifying the → *pharmacophore* for a series of compounds having the same biological activity; the pharmacophore here means a set of common structural and electronic features in active compounds [Dimoglo, 1985; Dimoglo, Bersuker *et al.*, 1988; Bersuker and Dimoglo, 1991; Bersuker, Dimoglo *et al.*, 1991].

The **Electron-Conformational method** (or **EC method**) is an extension of the Electronic-Topological method, which treats more explicitly the role of different conformations of molecules in determining activity. Moreover, this method is aimed at quantitatively predicting the biological activity on the basis of the presence of the pharmacophore, also accounting for the influence of pharmacophore flexibility and the concept of Anti-Pharmacophore Shielding (APS) [Bersuker, Bahçeci *et al.*, 1999b, 1999a, 2000b, 2000a; Bersuker, 2003].

Both methods are based on the same matrix representation of molecules, which is calculated from conformational geometries and → *quantum-chemical descriptors* of atoms and bonds. This

matrix is called **Electronic-Topological Matrix of Conjunction** (ETMC) in the framework of the ET method and **Electron-Conformational Matrix of Congruity** (ECMC) in the framework of the EC method. The matrix is symmetric of dimension $(A \times A)$, A being the number of atoms, and is constructed in the following way:

(a) the diagonal elements are electronic atomic parameters (often, atomic charges, HOMO and LUMO energies, values of the atomic $\rightarrow$ *Interaction Index*);

(b) the off-diagonal elements corresponding to bonded atoms are bond properties such as $\rightarrow$ *bond order*, bond energy, $\rightarrow$ *Wiberg index*, or polarizability;

(c) the off-diagonal elements corresponding to pairs of nonbonded atoms are their interatomic $\rightarrow$ *geometric distances* r_{ij}.

For a given conformation of a molecule the interatomic distances are fixed while the electronic parameters relative to atoms and bonds can be combined in different ways, each giving a different matrix. If all these matrices are taken together, a three-dimensional matrix is obtained with $A \times A \times K$ elements where A is the number of atoms and K is the number of all the considered combinations of electronic parameters (Figure E4).

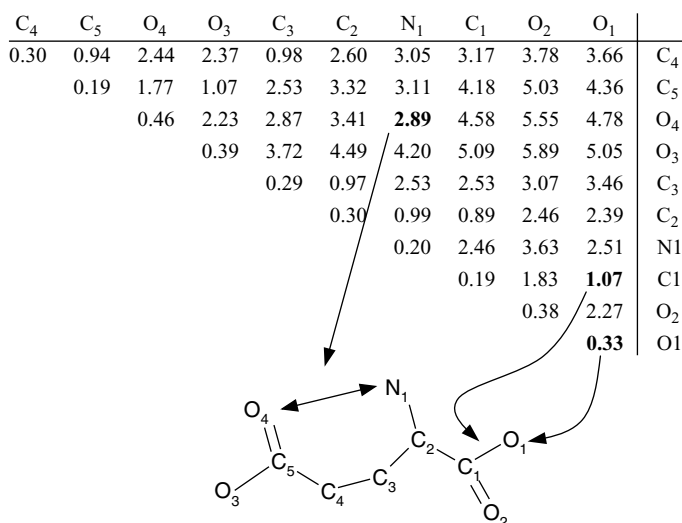| C_4	C_5	O_4	O_3	C_3	C_2	N_1	C_1	O_2	O_1	
0.30	0.94	2.44	2.37	0.98	2.60	3.05	3.17	3.78	3.66	C_4
	0.19	1.77	1.07	2.53	3.32	3.11	4.18	5.03	4.36	C_5
		0.46	2.23	2.87	3.41	**2.89**	4.58	5.55	4.78	O_4
			0.39	3.72	4.49	4.20	5.09	5.89	5.05	O_3
				0.29	0.97	2.53	2.53	3.07	3.46	C_3
					0.30	0.99	0.89	2.46	2.39	C_2
						0.20	2.46	3.63	2.51	N1
							0.19	1.83	**1.07**	C1
								0.38	2.27	O_2
									0.33	O1

Figure E4 The ECMC for glutamic acid. Interaction Index = 0.33; bond order = 1.07; interatomic distance = 2.89 Å.

The **ET method** allows to select molecular fragments that represent valuable information to design new active compounds [Güzel, 1996; Güzel, Saripinar *et al.*, 1997; Terletskaya, Shvets *et al.*, 1999; Shvets, Terletskaya *et al.*, 1999; Dimoglo, Shvets *et al.*, 2001; Altun, Kumru *et al.*, 2001].

To this end a four-step procedure is adopted.

(1) The Electronic-Topological Matrix of Conjunction is calculated for each molecule in the data set. If a molecule is found in a few stable conformations, all of them are treated as a

separate molecule and enter the analysis. After processing the ETMCs, only the conformation containing the pharmacophore is considered as the active one.

(2) Then, one of the most active compounds is chosen as the reference molecule and its ETMC (the template) is compared with all other ETMCs. By this comparison those matrix elements that are present in all active compounds but are absent in the inactive ones (i.e., active features or pharmacophore) are derived and represented by the **Electronic-Topological Submatrix of Conjunction** (ETSC).

(3) To derive this submatrix, which collects structural and electronic features responsible for activity, the flexibility of molecules is taken into account by choosing some tolerance limits for variation of diagonal (Δ_1) and off-diagonal (Δ_2 for bonded atoms and Δ_3 for nonbonded atoms) elements. Then, to decide which features are responsible for activity, two probabilistic functions are used:

$$(1) \ \ p_A(F_i) = \frac{n_1 + 1}{n_1 + n_3 + 2} \ \ \text{and} \ \ (2) \ \ p_A(F_i) = \frac{n_1 \cdot n_4 - n_2 \cdot n_3}{\sqrt{(n_1 + n_2) \cdot (n_1 + n_3) \cdot (n_2 + n_4) \cdot (n_3 + n_4)}}$$

where n_1 and n_2 are the numbers of molecules possessing and not possessing the feature F_i in the class of active compounds, respectively; n_3 and n_4 are the numbers of molecules possessing and not possessing the feature F_i in the class of nonactive compounds, respectively. The first function is similar to the $\rightarrow$ *Tversky similarity measure* with $\alpha = 1$ and $\beta = 0$ and can be interpreted as the fraction of active molecules possessing the activity feature F_i. The second function is the $\rightarrow$ *Pearson similarity coefficient*, which is related to both active and nonactive compounds.

(4) To verify the stability of the selected features the procedure is generally repeated using some different reference molecule.

The same procedure can be also used to select inactivity features by choosing one of the most inactive compounds as the reference molecule.

In the **EC method**, to allow the identification of a proper pharmacophore, first data of conformational analysis and electronic structure need to be calculated for every molecule in the training set and on the basis of these data the ECMC matrices are calculated for all the possible conformations of all the compounds.

By comparing the ECMC matrices of all the active compounds with those of the inactive ones, the **Electron-Conformational Submatrix of Activity** (ECSA) is derived, which contains those matrix elements that, within the chosen tolerance, are the same for all the active compounds and are absent in the same combination from the inactive ones. This EC Submatrix of Activity represents the pharmacophore and is equivalent to the Electronic-Topological Submatrix of Conjunction used in the ET method. The information contained in this matrix allows the designing of new active compounds as well as the screening of several compounds with respect to their activity. However, the presence of the pharmacophore is a very important necessary condition of activity, but it may not be sufficient for practical prediction of activity, because there may be other atoms and/or atomic groups that are positioned outside the pharmacophore and influence the molecule activity. These groups are divided into Anti-Pharmacophore Shielding (APS) groups defined as groups of atoms and charges outside the pharmacophore, which hinder

the proper ligand–receptor docking, diminishing the activity partially or completely, and other Auxiliary Groups (AG) that influence the activity in other way (e.g., groups responsible for hydrophobicity). The influence of the APS/AG groups on the activity is accounted for by introducing in the final model for activity some specific structural and electronic parameters, which need to be optimized by statistical analysis. Therefore, the model for activity (A) prediction, based on the Boltzmann distribution of each conformation of drug molecules and a function S for the energy difference in the ligand–receptor-binding interaction, is the following:

$$A_i = A_0 \cdot \frac{\sum_{m=1}^{M_i} e^{-E_{im}/kT} \cdot e^{-S_{im}} \cdot \delta_{im}(Pha)}{\sum_{m=1}^{M_i} e^{-E_{im}/kT}}$$

where A_i is the activity of the ith molecule, A_0 is a constant, M_i is the number of conformations of the ith molecule, kT the Boltzamann term, E_{im} the energy of the mth conformation of the ith molecule, and δ is the Dirac delta function equal to 1 when the pharmacophore (Pha) is present and zero otherwise. The function S_{im} accounts for the influence of APS/AG groups on the molecule activity and is defined as

$$S_{im} \equiv \frac{E'_{im}}{kT} = \sum_{j=1}^{N} b_j \cdot x_{im,j}$$

where E' is the contribution of APS/AG groups and pharmacophore flexibility to the ligand–receptor interaction energy, $x_{im,j}$ is the parameter that describes the jth APS/AG group in the mth conformation of the ith molecule, N is the number of chosen groups, and b are the regression coefficients to be estimated. The choice of $x_{im,j}$ parameters depends on the given data set. The parameter that describes the pharmacophore flexibility may be taken as the atomic → *Interaction Index* and interatomic distance between the pharmacophore atoms. The parameters for APS groups may be taken as the geometrical steric factors determined by their outstanding position with regard to the pharmacophore main plane and the basic skeleton.

A simplified model for activity is obtained by considering only the lowest energy conformation where the pharmacophore is present (P) and using a reference molecule (R) for which the activity A_R is known to determine the A_0 constant:

$$A_i = A_R \cdot e^{-(E_{iP} - E_{RP})/kT} \cdot e^{-(S_{iP} - S_{RP})}$$

where S_{iP} and S_{RP} are linear combinations of the chosen parameters x for APS/AG groups and pharmacophore flexibility, as described above, for the lowest energy conformation with Pha of the ith molecule and the reference molecule, respectively.

The **Electron-Conformational Approach** (ECA) is based on the same procedure as the ET method [Chumakov, Terletskaya *et al.*, 2000]. The main difference is the representation of molecules, which, in the framework of ECA, is defined in terms of a Set of Electronic and Conformational Parameters (SECPs) in place of the Electronic-Topological Matrix of Conjunction. Then, after choosing a compound as the reference, its SECPs are compared with the SECPs

of all other compounds to select the set of parameters that are common to all active compounds (i.e., activity features). The electronic → *quantum-chemical descriptors* are the dipole moment of the molecule, HOMO and LUMO energies, the energies of HOMO minus one and minus two orbitals (HOMO-1 and HOMO-2), their corresponding gap energies, atomic orbital coefficients and the derived reactivity indices, atomic charges, and nonpolar water-accessible area of the van der Waals molecular surface.

The conformational parameters are distances and torsion angles between pseudoatoms (PAs) and chain's atoms as well as dihedral angles between PAs. Pseudoatoms are used to replace flat fragments in the molecular graph, while the remaining atoms of a molecule are referred to as chains.

- ➢ **Electronic-Topological Submatrix of Conjunction** → Electronic-Topological method
- ➢ **electron-ion interaction potential** → average quasivalence number
- ➢ **electron isodensity contour surface** → molecular surface
- ➢ **electron-transition stochastic matrix** → MARCH-INSIDE descriptors
- ➢ **electrophilic atomic frontier electron density** → quantum-chemical descriptors
- ➢ **electrophilic charge** → quantum-chemical descriptors (⊙ electrophilic atomic frontier electron density)
- ➢ **electrophilic frontier electron density index** → quantum-chemical descriptors (⊙ electrophilic atomic frontier electron density)
- ➢ **electrophilic indices** → reactivity indices
- ➢ **electrophilicity index** → quantum-chemical descriptors (⊙ hardness indices)
- ➢ **electrophilic substituent constant** → electronic substituent constants (⊙ resonance electronic constants)
- ➢ **electrophilic superdelocalizability** → quantum-chemical descriptors
- ➢ **electropositivity of an atom** → substructure descriptors (⊙ pharmacophore-based descriptors)

◼ **electropy index** (ε)

It is an information index proposed to characterize the global electronic structure of molecules calculated from the molecular structure but avoiding quantum chemical approaches [Mekenyan, Bonchev et al., 1987]. It is defined as

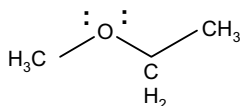$$\varepsilon = \log_2(N_{el}!) - \sum_{g=1}^{G} \log_2(n_g!)$$

where N_{el} is the total number of electrons in the molecule, that is, the sum of all electrons (inner and valence electrons) of all atoms; n_g is the number of electrons in the gth molecular subspace and G is the number of all possible subspaces.

Molecular subspaces are (a) the core parts of each atom type (e.g., C_{1S}, O_{1S}, N_{1S},); (b) σ and π bond spaces, where σ space is further divided into different independent bond spaces such as $C-C$, $C-H$, CH_2, CH_3, $C-O$, etc.; (c) the lone pairs on each atom also constitute an independent subspace.

The electropy index may be viewed as a measure of the degree of freedom for electrons to occupy different subspaces during the process of molecular formation.

Example E4

Calculation of the electropy index for methylethyl ether.

Molecular subspace type	Subspace electrons, n_g	Subspace of each type, n
1s(C)	2	3
1s(O)	2	1
CH_3	6	2
CH_2	4	1
C-O	2	2
C-C	2	1
oxygen lone pairs	4	1

no. of electrons, $N_{el} = 34$
no. of subspaces, $G = 11$

$$\varepsilon = \log_2(34!) - 3 \times \log_2(2!) - 1 \times \log_2(2!) + \\ -2 \times \log_2(6!) - 1 \times \log_2(4!) - 2 \times \log_2(2!) + \\ -1 \times \log_2(2!) - 1 \times \log_2(4!) = 92.642$$

➢ **electrostatic balance term** → GIPF approach
➢ **electrostatic factor** → electric polarization descriptors
➢ **electrostatic hydrogen-bond acidity** → Theoretical Linear Solvation Energy Relationships
➢ **electrostatic hydrogen-bond basicity** → Theoretical Linear Solvation Energy Relationships
➢ **electrostatic interaction fields** → molecular interaction fields
➢ **electrotopological descriptor** → charge descriptors (⊙ total absolute atomic charge)
➢ **electrotopological state index** ≡ *E-state index* → electrotopological state indices

■ **electrotopological state indices**
The electrotopological state S_i of the ith atom in the molecule, called **E-state index** (or **electrotopological state index**) gives information related to the electronic and topological state of the atom in the molecule and is defined as [Kier and Hall, 1990a, 1999b; Ivanciuc, 2008]:

$$S_i = I_i + \Delta I_i = I_i + \sum_{j=1}^{A} \frac{I_i - I_j}{(d_{ij} + 1)^k}$$

where I_i is the **intrinsic state** of the ith atom and ΔI_i is the field effect on the ith atom calculated as perturbation of the intrinsic state of ith atom by all other atoms in the molecule, d_{ij} is the → *topological distance* between the ith and the jth atoms, and A is the number of atoms. The exponent k is a parameter to modify the influence of distant or nearby atoms for particular studies. Usually it is taken as $k = 2$.

The intrinsic state of the ith atom is calculated by

$$I_i = \frac{(2/L_i)^2 \cdot \delta_i^v + 1}{\delta_i}$$

where L_i is the principal quantum number (2 for C, N, O, F atoms, 3 for Si, S, Cl, ...), δ_i^v is the number of valence electrons (→ *valence vertex degree*) and δ_i is the number of sigma electrons (→ *vertex degree*) of ith atom in the → *H-depleted molecular graph* (Table E11).

Table E11 Kier–Hall atom types.

No.	Atom group	Z	δ^v	δ	VSI	I_{AR}	I	KHE	Symbol
1	$-CH_3$	6	1	1	2	0	2.000	0.00	sCH3
2	$=CH_2$	6	2	1	3	0	3.000	0.25	dCH2
3	$-CH_2-$	6	2	2	4	0	1.500	0.00	ssCH2
4	$\equiv CH$	6	3	1	4	0	4.000	0.25	tCH
5	$=CH-$	6	3	2	5	0	2.000	0.25	dsCH
6	aCHa	6	3	2	5	1	2.000	0.25	aaCH
7	$>CH-$	6	3	3	6	0	1.333	0.00	sssCH
8	$=C=$	6	4	2	6	0	2.500	0.50	ddC
9	$\equiv C-$	6	4	2	6	0	2.500	0.50	tsC
10	$=C<$	6	4	3	7	0	1.667	0.25	dssC
11	aCa$-$	6	4	3	7	1	1.667	0.25	aasC
12	aaCa	6	4	3	7	1	1.667	0.25	aaaC
13	$>C<$	6	4	4	8	0	1.250	0.00	ssssC
14	$-NH_3[+1]$	7	2	1	3	0	3.000	0.25	sNH3
15	$-NH_2$	7	3	1	4	0	4.000	0.50	sNH2
16	$-NH_2-[+1]$	7	3	2	5	0	2.000	0.25	ssNH2
17	$=NH$	7	4	1	5	0	5.000	0.75	dNH
18	$-NH-$	7	4	2	6	0	2.500	0.50	ssNH
19	aNHa	7	4	2	6	1	2.500	0.50	aaNH
20	$\equiv N$	7	5	1	6	0	6.000	1.00	tN
21	$>NH-[+1]$	7	4	3	7	0	1.667	0.25	sssNH
22	$=N-$	7	5	2	7	0	3.000	0.75	dsN
23	aNa	7	5	2	7	1	3.000	0.75	aaN
24	$>N-$	7	5	3	8	0	2.000	0.50	sssN
25	$-N\lll$	7	5	3	8	0	2.000	0.50	ddsN (nitro)
26	aaNs	7	5	3	8	1	2.000	0.50	aasN (N-oxide)
27	$>N<[+1]$	7	5	4	9	0	1.500	0.25	ssssN (onium)
28	$-OH$	8	5	1	6	0	6.000	1.00	sOH
29	$=O$	8	6	1	7	0	7.000	1.25	dO
30	$-O-$	8	6	2	8	0	3.500	1.00	ssO
31	aOa	8	6	2	8	1	3.500	1.00	aaO
32	$-F$	9	7	1	8	0	8.000	1.50	sF
33	$-PH_2$	15	3	1	4	0	2.333	0.22	sPH2
34	$-PH-$	15	4	2	6	0	1.388	0.22	ssPH
35	$>P-$	15	5	3	8	0	1.073	0.22	sssP
36	$->P=$	15	5	4	9	0	0.806	0.11	dsssP
37	$->P<$	15	5	5	10	0	0.644	0.00	sssssP
38	$-SH$	16	5	1	6	0	3.222	0.44	sSH
39	$=S$	16	6	1	7	0	3.667	0.55	dS
40	$-S-$	16	6	2	8	0	1.833	0.44	ssS
41	aSa	16	6	2	8	1	1.833	0.44	aaS
42	$>S=$	16	6	3	9	0	1.221	0.33	dssS (sulfone)
43	$=>S=$	16	6	4	10	0	0.916	0.22	ddssS (sulfate)
44	$->S<-$	16	6	6	12	0	0.611	0.00	ssssssS
45	$-Cl$	17	7	1	8	0	4.111	0.67	sCl
46	$-$she	34	5	1	6	0	2.250	0.25	sSeH
47	$=Se$	34	6	1	7	0	2.500	0.31	dSe

(*Continued*)

Table E11 (*Continued*)

No.	Atom group	Z	δ^v	δ	VSI	I_AR	I	KHE	Symbol
48	—Se—	34	6	2	8	0	1.250	0.25	ssSe
49	>Se=	34	6	3	9	0	0.833	0.19	dssSe
50	>Se≪	34	6	4	10	0	0.625	0.13	ddssSe
51	—Br	35	7	1	8	0	2.750	0.38	sBr
52	—I	53	7	1	8	0	2.120	0.24	sI

Z, atomic number; δ^v, valence vertex degree; δ, vertex degree; *VSI*, valence state indicator; I_{AR}, aromatic indicator; *I*, intrinsic state; *KHE*, Kier–Hall electronegativity. Data from [Kier and Hall, 1999b].

The intrinsic state of an atom can be simply thought of as the ratio of π and lone-pair electrons over the count of the σ bonds in the molecular graph for the atom considered. Therefore, the intrinsic state reflects the possible partitioning of non-σ electron's influence along the paths starting from the considered atom; the less partitioning of the electron influence, the more available are the valence electrons for intermolecular interactions. The sum of the intrinsic states of all of the atoms is a molecular descriptor called **intrinsic state sum**; moreover, from the intrinsic state sum the → Q *polarity index* was derived.

The perturbation $\Delta_{ij} = I_i - I_j$ of the *i*th intrinsic state by the *j*th atom can be viewed as an "electronegative gradient" whose sign gives the direction of influence of surrounding atom intrinsic states.

From the definition of intrinsic states and field effects, it can be seen that large positive values of *E*-states S_i relate to atoms of high electronegativity and/or terminal atoms or atoms that lie on the mantle of the molecule; small or negative *E*-state values correspond to atoms possessing only σ electrons and/or buried in the interior of the molecule or close to higher electronegative atoms. Therefore, the *E*-state index is a measure of the electronic accessibility of an atom and can be interpreted as a probability of interaction with another molecule. However, the index cannot be considered a pure electronic descriptor: it is, in fact, a descriptor of atom polarity and steric accessibility.

Note that

$$\sum_{i=1}^{A} \Delta I_i = 0 \quad \rightarrow \quad \sum_{i=1}^{A} S_i = \sum_{i=1}^{A} I_i$$

the sum of the field effects over all atoms in the molecule being equal to zero.

This corresponds to an electronegativity equalization principle and means that the sum of the *E*-states in the molecule depends only on the number and type of atoms, not on their mutual interactions.

The electrotopological states are → *local vertex invariants*. After rescaling, they are also used as atomic weighting factors for the calculation of the → *WHIM descriptors*.

Since the *E*-state values derive from an H-depleted graph, they are calculated for each → *hydride group*, that is, they encode electronic and topological information about both heavy atoms and their bonded hydrogens. For molecules with high polar groups, these two contributions can be treated separately by the **hydrogen electrotopological state index** HS_i (or **HE-state index**), which was defined to complement the *E*-state index to encode electronic and topological

information about the hydrogens. It is defined as [Kier and Hall, 1999b]:

$$HS_i = KHE_i + [KHE_i - KHE(\mathrm{H}_i)] + \sum_{j \neq i} \frac{KHE_j - KHE(\mathrm{H}_i)}{(d_{ij} + 1)^2}$$
$$= KHE_i + [KHE_i + 0.2] + \sum_{j \neq i} \frac{KHE_j + 0.2}{(d_{ij} + 1)^2}$$

where KHE_i is the $\rightarrow$ *Kier–Hall electronegativity* of the ith heavy atom in the H-depleted molecular graph chosen as a measure of the intrinsic state of the attached hydrogen atom whose electronegativity $KHE(\mathrm{H}_i)$ is taken to be -0.2; the perturbation term is given by the sum over all other heavy atoms in the molecule of the difference between their electronegativity and hydrogen electronegativity divided by the square of the topological distance d; note that the distance is calculated between each j heavy atom and the ith heavy atom to which the hydrogen is bonded.

Therefore, given a H-depleted molecular graph two state values can be calculated for each vertex, the E-state value, which measures the electron density and accessibility of the atom, and the HE-state value, which measures the reaction and interaction ability of the bonded hydrogens, that is, the polarity of X$-$H bonds. Obviously, the HS values for atoms with no attached hydrogens are always equal to zero.

In this approach to the calculation of hydrogen E-states HS_i, topology is not considered a relevant factor in determining the E-state of a hydrogen, only the relative electronegativity is used to characterize the polarity of bonds with hydrogen in the molecule. However, in other definitions of hydrogen electrotopological states, the topology is also accounted for. In particular, in the first approach to HE-states the state of the hydrogen atom in the X$-$H bond is mainly determined by the electronegativity of the X atom and, to a lesser extent, by its topology [Kellogg, Kier *et al.*, 1996]. It is defined as

$$HS_i = I(\mathrm{H}_i) + \sum_j \frac{\Delta I_{ij}}{(d_{ij} + 1)^2} = \frac{(\delta_i^v - \delta_i)^2}{\delta_i} + \sum_j \frac{\Delta I_{ij}}{(d_{ij} + 1)^2}$$

where $I(\mathrm{H}_i)$ is the intrinsic state of the hydrogen bonded to the ith atom, defined in terms of square electronegativity of the ith atom to accentuate the electronic influence on H in the X$-$H bond. ΔI_{ij} is the perturbation of the hydrogen intrinsic state by the intrinsic state of the jth atom in the molecule and d is the topological distance. Hydrogen intrinsic states seem to be a measure of the H-donor ability of X$-$H groups.

Another approach [Kier and Hall, 1999b] is to apply the E-state definition for each vertex of the molecular graph where the hydrogens of polar groups ($-$OH, $-$NH, $-$COOH, etc.) are explicitly considered as independent vertices. In this case, the intrinsic state of polar hydrogens will be

$$I_i(\mathrm{H}) = \frac{(2/\mathrm{L}_i)^2 \cdot \delta_i^v + 1}{\delta_i} = 5$$

where L, δ^v, and δ always equal one. For each polar group, the difference between the HE-state and the E-state of the bonded atoms H$-$X reflects the polarity of the bond considered.

Based on the same approach used to define E-state indices, a **bond E-state index** BS_b was also tentatively proposed as [Kier and Hall, 1999b]

$$BS_b = BI_b + \Delta BI_{bt} = \sqrt{(I_i \cdot I_j)_b} + \sum_{t \neq b} \frac{BI_b - BI_t}{(d_{bt} + 1)^2}$$

where b is the considered bond formed by the atoms i and j, t runs over all the remaining bonds different from the bond b, BI is the bond intrinsic state defined by the intrinsic states I of the adjacent vertices, ΔBI the perturbation term, d_{bt} the → *topological edge distance* between bonds b and t. An alternative expression for the bond E-state index was proposed as [Tetko, Tanchuk *et al.*, 2001c]

$$BES_b = BI_b + \Delta BI_{bt} = \frac{(I_i + I_j)_b}{2} + \sum_{t \neq b} \frac{BI_b - BI_t}{(\bar{r}_{bt} + 1)^2}$$

where $\bar{r}_{bt}$ is the average bond length of the bonds b and t.

E-state and HE-state values were also used as atomic properties to calculate → *molecular interaction fields* [Kellogg, Kier *et al.*, 1996]. The **E-state fields** are defined by superimposing a 3D fixed grid over the molecule and calculating at each kth grid point an interaction energy value:

$$E_k = \sum_i S_i \cdot f(r_{ik})$$

where S_i is the electrotopological state value of the ith atom and $f(r_{ik})$ is a function of the distance r between each ith atom of the target molecule and the considered kth grid point. This function is not defined *a priori* but has to be empirically searched for; explored functions were $1/r$, $1/r^2$, $1/r^3$, $1/r^4$, and e^{-r}. Since the $1/r^n$ functions are discontinuous at grid points close to atoms, field default values of zero are set for grid points within the van der Waals envelope of the molecule. In the same way **HE-state fields** are calculated by using hydrogen electrotopological state values HS for each molecule atom instead of S values.

Moreover, **atom-type E-state indices** were proposed as molecular descriptors encoding topological and electronic information related to particular atom types in the molecule [Hall and Kier, 1995; Hall, Kier *et al.*, 1995]. They are calculated by summing the E-state values of all atoms of the same atom-type in the molecule or, alternatively, as average of the E-state values. Each atom-type is first defined by atom identity, based on the atomic number Z, and valence state, itself identified by the **valence state indicator** (*VSI*) defined as

$$VSI = \delta^v + \delta$$

where δ^v and δ are the → *valence vertex degree* and the → *vertex degree* of the atom; an aromatic indicator variable I_{AR} is also used to discriminate atoms of an aromatic system ($I_{AR} = 1$) from nonaromatic atoms ($I_{AR} = 0$). To distinguish particular atoms, classified as the same atom type according to atom number, valence state indicator and aromatic indicator, the analysis of bonded atoms and the difference between valence δ^v and simple vertex degree δ are used. The symbol of each atom-type E-state index is a composite of three parts (Table E11). The first part is "S" which refers to the sum of the E-states of all atoms of the same type. The second part is a string representing the bond types associated with the atom ("s", "d", "t", "a" for single, double, triple, and aromatic bonds, respectively). The third part is the symbol identifying the chemical element and eventual bonded hydrogens, such as CH_3, CH_2, F, and so on.

The atom-type E-state indices combine structural information about the electron accessibility associated with each atom type, an indication of the presence or absence of a given atom type and a count of the number of atoms of a given atom type.

Atom-type E-state counts were proposed as simple counts of the E-state atom types in a molecule [Butina, 2004].

Moreover, **atom-type *HE*-state indices** were proposed as molecular descriptors calculated by summing hydrogen electrotopological states of all atoms of the same atom-type [Kier and Hall, 1999b]. **Bond-type *E*-state indices** were analogously defined by summing bond *E*-state index values of all the edges of the same type [Tetko, Tanchuk *et al.*, 2001c].

A → *Balaban-like index*, called → *E-state topological parameter*, was derived from *E*-state indices.

Two sets of molecular descriptors, called **intrinsic state pseudoconnectivity indices** ψ, one based on the intrinsic states *I* and the other on the scaled electrotopological state values *E*, were proposed as [Pogliani, 2000b, 2004]:

1. $\qquad {}^{S}\psi_I = \sum_{i=1}^{A} I_i \qquad\qquad\qquad\qquad {}^{S}\psi_E = \sum_{i=1}^{A} E_i$

2. $\qquad {}^{0}\psi_I = \sum_{i=1}^{A} (I_i)^{-1/2} \qquad\qquad\qquad {}^{0}\psi_E = \sum_{i=1}^{A} (E_i)^{-1/2}$

3. $\qquad {}^{1}\psi_I = \sum_{i=1}^{A-1}\sum_{j=i+1}^{A} a_{ij} \cdot (I_i \cdot I_j)^{-1/2} \qquad {}^{1}\psi_E = \sum_{i=1}^{A-1}\sum_{j=i+1}^{A} a_{ij} \cdot (E_i \cdot E_j)^{-1/2}$

4. $\qquad {}^{T}\psi_I = \prod_{i=1}^{A} (I_i)^{-1/2} \qquad\qquad\qquad {}^{T}\psi_E = \prod_{i=1}^{A} (E_i)^{-1/2}$

5. $\qquad {}^{0}\psi_{Id} = (-0.5)^{A} \cdot \prod_{i=1}^{A} I_i \qquad\qquad {}^{1}\psi_{Ed} = (-0.5)^{A} \cdot \prod_{i=1}^{A} E_i$

6. $\qquad {}^{1}\psi_{Id} = (-0.5)^{(A+C-1)} \cdot \prod_{b}(I_i + I_j)_b \quad {}^{1}\psi_{Ed} = (-0.5)^{(A+C-1)} \cdot \prod_{b}(E_i + E_j)_b$

7. $\qquad {}^{1}\psi_{Is} = \prod_{b}(I_i + I_j)_b^{-1/2} \qquad\qquad {}^{1}\psi_{Es} = \prod_{b}(E_i + E_j)_b^{-1/2}$

where *A* is the total number of atoms; in the equations 1, 2, 4, and 5 the summation/product runs over all the atoms, while, in equation 3, it is over all the atoms, but only pairs of bonded atoms give contributions different from zero, a_{ij} being the elements of the → *adjacency matrix*. In equation 6, *C* is the → *cyclomatic number*; in equations 6 and 7, only contributions from pairs of bonded atoms are considered.

The pseudoconnectivity descriptors Ψ_E are calculated from the electrotopological *E*-state values S_i transformed to avoid negative *S* values. Then, scaled electrotopological values *E* are calculated as $E_i = S_i + 5.5$, where -5.5 is the *S* value for the carbon atom in CF_4, which is the lowest *S* value a carbon atom can assume. These descriptors were used to model melting point and crystal density of aminoacids and alkanes, enthalpies of metal halides, as well as some biological activities [Pogliani, 2001a, 2002a, 2002b, 2002c, 2003a, 2003b, 2005a, 2005b, 2006b].

Other descriptors based on a modification of intrinsic and *E*-states are → *MEDV descriptors*.

📖 Additional references are listed in the thematic bibliography (see Introduction).

- ➢ **elongation** → graph
- ➢ **elongation** → grid-based QSAR techniques (⊙ VolSurf descriptors)
- ➢ **elongation/elongation-fixed ratio** → grid-based QSAR techniques (⊙ VolSurf descriptors)
- ➢ **embedded cluster graph** → biodescriptors (⊙ proteomics maps)
- ➢ **embedded correlation** → statistical indices (⊙ correlation measures)
- ➢ **embedded graph of partial order** → biodescriptors (⊙ proteomics maps)
- ➢ **embedded neighborhood graph** → biodescriptors (⊙ proteomics maps)
- ➢ **embedded zigzag curve** → biodescriptors (⊙ proteomics maps)
- ➢ **embedding frequencies** → cluster expansion of chemical graphs

◼ **empirical descriptors**

The class of the empirical descriptors is a fuzzy, not well-defined class. In principle, empirical descriptors are those not defined on the basis of a general theory such as, for example, quantum chemistry or graph theory. Rather they are defined by practical rules derived from chemical experience, thus considering specific or local structural factors present in the molecules, often sets of congeneric compounds. As a consequence, in most of the cases, empirical descriptors represent limited subsets of compounds and cannot be extended to classes of compounds different from those for which they were defined. Empirical descriptors have not to be confused with experimentally derived descriptors even if it is well known that several of them are empirically derived.

Empirical descriptors can be user-defined values for discriminating among special molecular fragments or number of occurrences of local specific atom/fragment within a molecule.

An example of useless empirical descriptor is the **Sadhana index** (or *Sadna index*) [Khadikar, Agrawal *et al.*, 2002; Khadikar, Joshi *et al.*, 2004] defined for polyacene molecules as $Sd = 2$ $h \times (5h + 1)$, where h is the number of benzene units. This index is almost perfectly correlated to the → *PI index*, to the square count of benzene units, and any quadratic combination of the coefficients of h and h^2, such as $5h^2 + h$ or $20h^2 + 4h$.

Other examples of empirical descriptors are the → *Taillander index* (restricted to substituted benzenes), → *second-grade structural parameters* (restricted to alkenes), → *polar hydrogen factor* (restricted to halogenated hydrocarbons), → *hydrophobic fragmental constants*, → *six position number*, → *Idoux steric constant*, → *hydrophilicity index*, → *adsorbability index*, → *bond flexibility index*, and → *atomic solvation parameter*.

📖 [Carlton, 1998; Chiorboli, Piazza *et al.*, 1993c]

- ➢ **endospectral graph** → self-returning walk counts
- ➢ **endospectral vertices** → self-returning walk counts
- ➢ **end point mean square distance index** → distance matrix
- ➢ **end-to-end distance** → size descriptors
- ➢ **energy-based descriptors** → quantum-chemical descriptors
- ➢ **energy moments** → quantum-chemical descriptors
- ➢ **enthalpic fields** → molecular interaction fields
- ➢ **enthalpic hydrophobic substituent constant** → lipophilicity descriptors (⊙ Hansch–Fujita hydrophobic substituent constants)

> ➤ **enthalpies** → physico-chemical properties
> ➤ **entropic fields** → molecular interaction fields
> ➤ **entropic hydrophobic substituent constant** → lipophilicity descriptors (⊙ Hansch–Fujita hydrophobic substituent constants)

▨ **environmental indices**

In recent years, a great importance has been given to QSAR approaches for modeling and predicting the environmental behavior, fate, and toxicity of chemicals, that is, **environmental QSAR** [Dearden, 2002; Devillers and Karcher, 1991; Karcher and Devillers, 1990a, 1990b; Karcher and Karabunarliev, 1996; Sabljić, 1990]. Research has mainly focused on some chemical classes of environmental interest, such as Persistent Organic Pollutants (POPs), Volatile Organic Compounds (VOC), Hazardous Air Pollutants (HAP), Persistent-Bioaccumulative-Toxic (PBT) pollutants.

Numerical quantities, experimentally determined or estimated by statistical or computational approaches, which measure the environmental behavior, fate, and toxicity of chemicals, are molecular descriptors usually referred to as environmental indices. The → *octanol–water partition coefficient* (K_{ow}, log P) is the most well-known environmental index used as the measure of lipophilicity of compounds. Together with log P and the → *soil sorption partition coefficient* (K_{oc}) [Baker, Mihelcić *et al.*, 2001; Uddameri and Kuchanur, 2004], other quantities have to be considered as relevant for environmental studies. Some of these have been defined to describe mobility, biodegradability, bioaccumulation, metabolism, partition, and toxicity of chemicals, thus becoming relevant to human health and environmental safety assessment.

Some definitions of the most popular environmental descriptors are given below.

• **half lifetime** ($t_{1/2}$)

The half lifetime of a compound, subject to exponential decay, is the time required for the compound to decay to half of its initial value. Although this concept originated from the study of radioactive decay, it applies to many other fields as well, including phenomena that are described by nonexponential decays. It is mathematically defined as

$$t_{1/2} = \ln(2) \cdot \tau$$

where τ is the mean lifetime.

The half lifetime of a chemical is calculated as the length of time it takes for the concentration of that chemical to be reduced by one-half relative to its initial level, assuming first-order decay kinetics. It can be estimated for all major environmental compartments (water, air, soil, sediments, and biota).

The half lifetime in air is related to the atmospheric degradation ability of a chemical, measured by the rate constant of its reactions with free radicals (e.g., OH•, NO$_3^•$) and ozone O$_3$ or of photochemical reactions [Gramatica and Papa, 2007; Gramatica, Pilutti *et al.*, 2003b].

In water, soil, and sediments, besides physical and chemical reactions, the enzymatic biological activity, mainly due to microorganisms (biodegradation) plays a relevant role.

• **persistence**

The persistence of a chemical is the length of time it remains in a particular environment (water, soil, air, and sediment) in an unchanged form before it is physically transported to another

compartment and/or is chemically or biologically transformed [Gramatica, Consolaro *et al.*, 2001; Gramatica, Papa *et al.*, 2004; Leip and Lammel, 2004; Pavan and Worth, 2008]. The longer a chemical persists, the higher the potential for human or environmental exposure to it.

The experimental determination of the persistence is generally based on the **degradability** of the chemical, that is, a chemical that is degraded in an experimental test system is usually considered not persistent.

The degradation process is often characterized by the extent of degradation and the nature of the degradation process. Then, *primary degradation* refers to the production of organic derivatives that exhibit their own degradation properties; *mineralization* refers to the complete degradation of an organic chemical to stable inorganic species; *abiotic degradation* refers to transformations such as reduction, oxidation, hydrolysis, and photodegradation; *biodegradation* refers to transformation by enzymatic reactions in microorganisms [Pavan and Worth, 2008].

- **bioconcentration factor** (*BCF*)

The bioconcentration factor is the concentration of a chemical in a tissue per concentration of the chemical in water (generally adimensional) [Pavan, Netzeva *et al.*, 2008]. This physical property characterizes the uptake of pollutants due to chemical partitioning from environmental phase (e.g., air or water) into an organic phase (e. g., lipids or proteins) through an exchange surface (e.g., gills of fish).

A simple scale for *BCF* values is the following: high potential: $BCF > 1000$; moderate potential: $1000 > BCF > 250$; low potential: $BCF < 250$.

The bioconcentration factor for aquatic organisms is related to the octanol–water partition coefficient (K_{ow}) by the following equation:

$$\log(BCF) = a + b \cdot \log K_{ow}$$

where coefficients a and b depend upon the type of organism (e.g., fish, crustaceans).

Comparable equations can be developed for bioconcentration from air, by using K_{oa} (octanol–air partition coefficient).

   📖 [Bermúdez-Saldaña, Escuder-Gilabert *et al.*, 2005; Cheng, Kontogeorgis *et al.*, 2005; Dearden, 2002; Devillers, Domine *et al.*, 1998; Feng, Han *et al.*, 1996b; Govers, Rupert *et al.*, 1984; Gramatica, 2001; Gramatica and Papa, 2003, 2005; Ivanciuc, 1998b; Khadikar, Singh *et al.*, 2003; Papa, Dearden *et al.*, 2007; Roy, Sanyal *et al.*, 2006; Russom, Breton *et al.*, 2003; ; Sabljić, 1988; Sabljić and Protic, 1982a; Vighi, Gramatica *et al.*, 2001; Wei, Zhang *et al.*, 2001; Zhao, Yuan *et al.*, 1997]

- **bioaccumulation**

Bioaccumulation represents the uptake of a chemical through all routes of entry into the organism, in particular through food. Therefore, it is not a simple physical–chemical process, because active absorption in the digestive system may occur.

Biomagnification occurs when the concentration of a chemical increases substantially (often many orders of magnitude) at different levels of the food chain, from primary producers to top predators. Biomagnification is possible for chemicals that are very persistent in living organisms (neither metabolized nor excreted) and that can be efficiently stored in some tissues (usually lipids or proteins).

- **leaching indices**

They are environmental indices specifically proposed to study the environmental fate of pesticides.

The **GUS index** assesses the leachability of molecules and the possibility of finding these chemicals in groundwater [Gustafson, 1989; Papa, Castiglioni *et al.*, 2004]. It is defined as

$$\text{GUS} = \log_{10}(t_{1/2}) \cdot [4 - \log_{10}(K_{oc})]$$

where $t_{1/2}$ is the half lifetime (in days), quantifying the soil persistence, and K_{oc} is the organic carbon partition coefficient, quantifying the mobility in soil.

The **LEACH index** is a leaching index assessing the potential degree of groundwater and river water contamination [Laskowski, Goring *et al.*, 1982]. It is defined as

$$\text{LEACH} = \frac{S_w \cdot t_{1/2}}{V_p \cdot K_{oc}}$$

where S_w is the water solubility (mg/l), $t_{1/2}$ is the degradation half lifetime in soil (in days), V_p the vapour pressure (Pa), and K_{oc} is the organic carbon partition coefficient. The lower the LEACH value the lower the risk of contamination. LEACH values are expressed on a logarithmic scale to allow comparison with other indices. A modified version of the LEACH index, **modified LEACH index**, was also proposed without taking vapour pressure into account, to avoid a double counting of volatilization that is already considered in disappearance half lifetime:

$$\text{LEACH}_{\text{MOD}} = \frac{S_w \cdot t_{1/2}}{K_{oc}}$$

Based on the Principal Component Analysis, the **LIN index** (leaching index) and the **VIN index** (volatility index) were defined in terms of the first and second PCs, respectively, explaining 92.7% of the total variance [Gramatica and Di Guardo, 2002]. PCs were calculated on a data set of 135 pesticides, described by vapour pressure (V_p), Henry's law constant (H), water solubility (S_w), and octanol/water (K_{ow}) and organic carbon (K_{oc}) partition coefficients. The LIN and VIN indices are defined as the following:

$$\text{LIN} = -0.531 \cdot \log K_{ow} + 0.518 \cdot \log S_w - 0.495 \cdot \log K_{oc} - 0.023 \cdot \log V_p - 0.452 \cdot \log H$$
$$\text{VIN} = -0.034 \cdot \log K_{ow} + 0.211 \cdot \log S_w + 0.202 \cdot \log K_{oc} + 0.837 \cdot \log V_p + 0.461 \cdot \log H$$

The **Global Leachability Index (GLI index)** [Papa, Castiglioni *et al.*, 2004] was defined by Principal Component Analysis, condensing information derived from GUS index, modified LEACH index, and LIN index.

The first PC, explaining 87.5% of the total variance, was assumed as representing the risk (higher positive values higher the risk):

$$\text{GLI} = 0.579 \cdot \text{LIN} + 0.558 \cdot \text{GUS} + 0.595 \cdot \text{LEACH}_{\text{MOD}}$$

GLI values less than -0.5 indicate low-risk compounds, values between -0.5 and 1 medium-risk compounds, and values greater than 1 high-risk compounds.

- **Global Warming Potential** (GWP)

Global warming potential is a measure of how much a given mass (a ton) of greenhouse gas (GHG) is estimated to contribute to global warming, evaluated as its accumulated

radiative effect. It is a relative scale that compares the gas in question to that of the same mass of carbon dioxide CO_2 (whose GWP is by definition 1). A GWP is calculated over a specific time interval (usually 20, 100, or 500 years) [Ivanciuc and Ivanciuc, 2002].

Other properties of chemicals, encountered in environmental studies, are *Long Range Transport* (LRT) [Beyer, Mackey *et al.*, 2000; Leip and Lammel, 2004; Wania and Dugani, 2003], defined as the atmospheric transport of air pollutants within a moving air mass for a distance greater than 100 km, the *mobility index* [Gramatica, Papa *et al.*, 2004], and *Atmospheric Persistent Index* (ATPIN) to evaluate the atmospheric degradability of chemicals [Gramatica, Pilutti *et al.*, 2002, 2003b].

📖 [Basak and Mills, 2005; Cronin, Walker *et al.*, 2003; Drefahl and Reinhard, 1993; Govers, 1990; Halfon, Galassi *et al.*, 1996; Halfon and Reggiani, 1986; Jaworska, Comber *et al.*, 2003; Koch, 1982; Okouchi, Saegusa *et al.*, 1992; Rorije, Van Wezel *et al.*, 1995; Sabljić, 2001; Sabljić and Piver, 1992; Todeschini and Gramatica, 1997c]

➢ **environmental QSAR** → environmental indices
➢ **environment connectivity descriptors** → connectivity indices
➢ **E_K polarity scale** → Linear Solvation Energy Relationships (⊙ dipolarity/polarizability term)
➢ **E_T polarity scale** → Linear Solvation Energy Relationships (⊙ dipolarity/polarizability term)
➢ **equilibrium constants** → physico-chemical properties
➢ **equilibrium electronegativity** → electronegativity
➢ **equipoise random walks** → walk counts
➢ **equipotent walks** → walk counts

■ **equivalence classes**
Subsets of equivalent elements according to a specified equivalence relation.

A **decomposition** N of a system containing N elements is any partition of these elements into disjoint subsets of equivalent elements by a specified equivalence relation.

A **finite probability scheme** can be associated with this decomposition as the following:

|	
Equivalence classes	$1, 2, \ldots, G$
Element partition	$n_1, n_2, \ldots, n_g, \ldots, n_G$
Probability distribution	$p_1, p_2, \ldots, p_g, \ldots, p_G$

where G is the total number of equivalence classes, $p_g = n_g/N$ is the probability of a randomly chosen element belonging to the gth subset having n_g elements and

$$N = \sum_{g=1}^{G} n_g.$$

→ *Information content* is a fundamental measure derived from the partitioning of elements in equivalence classes; several molecular descriptors are derived as → *information indices*.

Example E5

Equivalence classes of 5-methyl-1,3,4-oxathiazol-2-one based on the atom types.
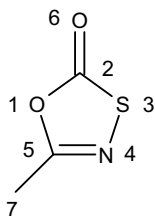
Atom	C	O	N	S
n	3	2	1	1
p	3/7	2/7	1/7	1/7

➤ **E$_R$ radical parameter** → electronic substituent constants
➤ **error rate** → classification parameters
➤ **error standard deviation** ≡ *residual standard deviation* → regression parameters
➤ **error sum of squares** ≡ *residual sum of squares* → regression parameters
➤ **ESSR** ≡ *Extended Set of Smallest Rings* → ring descriptors
➤ **E-state index** → electrotopological state indices
➤ **E-state fields** → electrotopological state indices
➤ **E-state topological parameter** → Balaban distance connectivity index
➤ **Estrada Generalized Topological Indices** → variable descriptors
➤ **Estrada index** → spectral indices (⊙ subgraph centrality)

◼ **ETA indices** (≡ *Extended Topochemical Atom indices*)
ETA indices [Roy and Ghosh, 2003] are both local vertex and graph invariants, defined in the framework of the **Valence Electron Mobile environment** (**VEM environment**) theory [Pal, Sengupta *et al.*, 1988; Pal, Sengupta *et al.*, 1989], according to which a vertex in the → *H-depleted molecular graph* is considered to be consisted of a core and a valence electronic environment.

The **core count** α_i is a local vertex invariant calculated as

$$\alpha_i = \frac{Z_i - Z_i^v}{Z_i^v} \cdot \frac{1}{L_i - 1}$$

where Z, Z^v, and L are the atomic number, the valence electron number, and the principal quantum number, respectively. The sum of the α values of the nonhydrogen atoms is a simple molecular descriptor related to molecular bulk.

By combining the core count α of an atom with its valence electron number Z^v, the **electronegativity ETA measure** ε_i for the ith atom was defined as

$$\varepsilon_i = -\alpha_i + 0.3 \cdot Z_i^v$$

The **VEM count** β_i is another local vertex invariant defined as

$$\beta_i = [\beta_s]_i + [\beta_{ns}]_i = \left[\sum_{\sigma(i)} f_\sigma\right] + \left[\sum_{\pi(i)} f_\pi + f_{LP(i)}\right]$$

where $[\beta_s]_i$ and $[\beta_{ns}]_i$ are two other local vertex invariants accounting for σ bonds and π bonds plus lone pairs of each ith atom, respectively, f_σ is the contribution of a σ bond and f_π the

contribution of a π bond, and f_{LP} a correction factor accounting for lone pair electrons; the two summations run over bonds formed by the ith atom.

The contribution of a sigma bond f_σ is equal to 0.5 for two bonded atoms of similar ETA electronegativity ($\Delta\varepsilon \leq 0.3$) and 0.75 for two bonded atoms of different electronegativity ($\Delta\varepsilon > 0.3$). The contribution of a π bond f_π arises only from multiple bonds (i.e. double and triple bonds) and is equal to (a) $f_\pi = 1$ for two bonded atoms of similar electronegativity ($\Delta\varepsilon \leq 0.3$); (b) $f_\pi = 1.5$ for two bonded atoms of different electronegativity ($\Delta\varepsilon > 0.3$) or for conjugated (nonaromatic) π systems; and (c) $f_\pi = 2$ for aromatic π systems. The term f_{LP} is equal to 0.5 per atom with a lone pair of electrons capable of resonance with aromatic ring (e.g., nitrogen of aniline, oxygen of phenol, etc.).

From the atomic VEM counts, two molecular descriptors are derived:

$$\beta_s = \frac{\sum_{i=1}^{A} [\beta_s]_i}{A} \qquad \beta_{ns} = \frac{\sum_{i=1}^{A} [\beta_{ns}]_i}{A}$$

where A is the number of nonhydrogen atoms; β_s accounts for all the σ bonds in the molecule and can be considered a relative measure of the number of electronegative atoms in the molecule, while β_{ns} is calculated from all π bonds and electron lone pairs and can be considered a relative measure of electron-richness (unsaturation) of the molecule.

From the two vertex invariants α and β, the **VEM vertex count** γ_i is defined as

$$\gamma_i = \frac{\alpha_i}{\beta_i}$$

The **composite ETA index** η is a molecular descriptor defined as

$$\eta = \sum_{i=1}^{A-1} \sum_{j=i+1}^{A} \left(\frac{\gamma_i \cdot \gamma_j}{d_{ij}^2} \right)^{1/2}$$

where d_{ij} is the topological distance between i–j atoms.

Defined analogously to the composite ETA index, but considering only contributions from pairs of bonded atoms, the **local ETA index** is calculated as

$$\eta^{loc} = \sum_{i=1}^{A-1} \sum_{j=i+1}^{A} a_{ij} \cdot (\gamma_i \cdot \gamma_j)^{1/2}$$

where a_{ij} are the elements of the $\rightarrow$ *adjacency matrix*, equal to 1 only for pairs of bonded atoms and zero otherwise.

Moreover, the **composite reference ETA index**, denoted as η_R, is calculated as the composite ETA index from a molecular graph where all heteroatoms are substituted by carbon atoms and all the multiple bonds by single bonds. Then, the **functionality index** η_F is defined as the difference between the composite reference ETA index and the composite ETA index:

$$\eta_F = \eta_R - \eta$$

This index was proposed to measure the molecule functionality, here intended as the presence of heteroatoms and multiple bonds. To avoid molecular size dependence, this functionality index is normalized by the number of atoms A:

$$\eta'_F = \frac{\eta_R - \eta}{A}$$

In a similar way to the functionality index, the **local functionality index** is defined as

$$\eta_F^{loc} = \eta_R^{loc} - \eta^{loc}$$

where η_R^{loc} is the local index for the corresponding reference alkane.

The **branching ETA index** η_B is calculated from the local ETA index of the reference alkane η_R^{loc} of the considered compound compared to the index value η_N^{loc} of the corresponding normal alkane, that is, the straight chain graph containing the same number of vertices as the compound considered:

$$\eta_B = \eta_N^{loc} - \eta_R^{loc} + 0.086 \cdot NRG$$

where *NRG* is the number of rings in the molecular graph of the reference alkane. η'_B is the branching ETA index normalized by the number of atoms *A*.

The calculation of η_N^{loc} can be easily performed by the following relationship:

$$\eta_N^{loc} = 1.414 + (A-3) \cdot 0.5$$

where *A* is the total number of nonhydrogen atoms and the relationship holds only for compounds for which $A > 2$.

The **shape ETA indices** are derived from the core counts α as

$$\frac{\left[\sum_i \alpha_i\right]_p}{\sum_{i=1}^{A} \alpha_i}, \qquad \frac{\left[\sum_i \alpha_i\right]_y}{\sum_{i=1}^{A} \alpha_i}, \qquad \frac{\left[\sum_i \alpha_i\right]_x}{\sum_{i=1}^{A} \alpha_i}$$

where p, y, and x stand for the summation of α values of the vertices that are joined to one, three, and four other nonhydrogen atoms, respectively.

The contribution of a specific *i*th atom to the composite ETA index was also defined as local vertex invariant and called **atom-level composite ETA index**:

$$[\eta]_i = \sum_{j=1}^{A} \left(\frac{\gamma_i \cdot \gamma_j}{d_{ij}}\right)^{1/2} \quad j \neq i$$

Analogously, the reference and functionality atom-level indices and their normalized counterparts were also derived and denoted as $[\eta_R]_i$, $[\eta_F]_i$, and $[\eta'_F]_i$.

ETA indices are an extension of the **TAU indices** (or *Topochemically Arrived Unique indices*) [Pal, Purkayastha *et al.*, 1992; Pal, Sengupta *et al.*, 1988, 1989, 1990], which were defined some years before in the framework of a previous version of the Valence Electron Mobile environment (VEM environment). TAU indices are calculated from previous definitions of core count and VEM count and include four indices: the composite topochemical index, denoted by *T* (similar to the composite ETA index), the functionality index, denoted by *F*, the skeletal index, denoted by T_R, and the simple branching index, denoted by *B*. In QSAR studies, these indices were used in combination with $\rightarrow$ *STIMS indices*, $\rightarrow$ *connectivity indices*, and some $\rightarrow$ *information indices* [Roy, Pal *et al.*, 1999, 2001].

📕 [Pal, Purkayastha *et al.*, 1992; Pal, Sengupta *et al.*, 1990; Roy and Ghosh, 2004a, Roy and Ghosh, 2004b, Roy and Ghosh, 2004c, Roy and Ghosh, 2005, Roy and Ghosh, 2006a, Roy and Ghosh, 2006b, Roy and Ghosh, 2007; Roy and Saha, 2003a, 2003b, 2004; Roy, Sanyal *et al.*, 2006, 2007; Roy and Sanyal, 2006; Roy and Toropov, 2005]

- ➤ **ET method** ≡ *Electronic-Topological method*
- ➤ **Euclidean-adjacency map matrix** → biodescriptors (⊙ proteomics maps)
- ➤ **Euclidean connectivity index** → connectivity indices
- ➤ **Euclidean degree** ≡ *geometric distance degree* → molecular geometry
- ➤ **Euclidean distance** → similarity/diversity (⊙ Table S7)
- ➤ **Euclidean distance matrix** → similarity/diversity
- ➤ **Euclidean-distance map matrix** → biodescriptors (⊙ proteomics maps)
- ➤ **Eulerian walk** → graph

■ **Euler formula**

The fundamental relation between the number of vertices V, edges E, and faces F of convex polyhedra was proposed in 1758 by Euler [Euler, 1758] as

$$V - E + F = 2$$

This relationship holds for any division of a sphere into polygons.

Euler's formula relating the number of edges, vertices, and faces of a convex polyhedron was studied and generalized by Cauchy [Cauchy, 1813] and [L'Huillier, 1861] and is at the origin of topology.

From these primary indices, that is, vertices, edges, and faces, two secondary topological indices were derived, called **Schläfli indices** [Bucknum and Castro, 2005a], namely, the polygonality (n) and the connectivity (p).

Polygonality refers to the weighted average number of sides of the polygonal faces of a polyhedron, computed by drawing polygons and also considering the terminal atoms. *Connectivity* refers to the average of connectivity of the vertices of a polyedron. Moreover, the ratio of polygonality to connectivity was proposed as a compactness index and called **Schläfli topological form index** [Bucknum and Castro, 2005b]:

$$I = \frac{n}{p}$$

Based on the two topological identities, (1) each edge of a polyhedron is shared by two faces, then $n \times F = 2 \times E$, and (2) each edge terminates with two vertices, then $p \times V = 2 \times E$, the following Schläfli relationship was found:

$$\frac{1}{n} + \frac{1}{p} - \frac{1}{2} = \frac{1}{E}$$

The Schläfli relation establishes a connection between secondary topological indices, n and p, and primary topological indices V, E, and F.

Poligonality and connectivity can be calculated not only for 3D pattern but also for 2D patterns, introducing the concept of cell. A cell is a topological concept, like a sphere, which involves the division of the plane into fused polygons in which some edges form a boundary.

In effect, the Euler relation in a 2D space, a cell, is

$$V - E + F = 1$$

Noting that faces are exactly what in graph theory are called cycles, the previous Euler formula provides the definition of $\rightarrow$ *cyclomatic number C*:

$$F \equiv C = E - V + 1 \equiv B - A + 1$$

where A and B are the number of vertices and edges in a $\rightarrow$ *molecular graph*.

■ **EVA descriptors** ($\equiv$ *EigenVAlue descriptors*)
EVA descriptors were proposed by Ferguson *et al.* [Ferguson, Heritage *et al.*, 1997; Turner, Willett *et al.*, 1997] as an approach to extract chemical structural information from mid- and near-infrared spectra. The approach is to use, as a multivariate descriptor, the vibrational frequencies of a molecule, a fundamental molecular property characterized reliably and easily from the potential energy function. The EigenVAlue (EVA) descriptor is a function of the eigenvalues obtained from the normal coordinate matrix; it corresponds to the fundamental vibrational frequencies of the molecule, which can be calculated using standard quantum or molecular mechanical methods from $\rightarrow$ *computational chemistry*.

The eigenvectors, corresponding to atomic displacement, are not considered as molecular descriptors.

Since the number of vibrational normal modes varies with the number of atoms A in a molecule (actually $3A-6$ for a molecule without axial symmetry or $3A-5$ for a linear molecule), each set of eigenvalues will generally be of different dimensionality. Thus, to obtain comparability among the molecules and $\rightarrow$ *uniform-length descriptors*, the frequency range chosen is 0 and $4000\,cm^{-1}$ to encompass the frequencies of all fundamental molecular vibrations, and the eigenvalues are projected onto a bounded frequency scale (BFS) where the vibrational frequencies are represented by points along this axis, obtaining a scale of fixed dimensionality. Then a Gaussian function of fixed standard deviation σ is centered at each eigenvalue projection over the BFS axis, resulting in a series of $3A-6$ (or $3A-5$) identical and overlapping Gaussians (Figure E5).

The value of the **EVA function** at any point x on the BFS axis is determined by summing the contributions from each and every one of the $3A-6$ (or $3A-5$) overlaid Gaussians at that point:

$$EVA_x = \sum_{i=1}^{3A-6} \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-(x-\lambda_i)^2/2\sigma^2}$$

where λ_i is the ith vibrational frequency (eigenvalue) for the molecular structure.

Finally, the EVA function is sampled at fixed increments of $L\,cm^{-1}$ along the BFS axis; this sampling results in $4000/L$ values that define the EVA uniform-length descriptor.

The choice of σ defines the degree to which the fundamental vibrations overlap: σ values determine the number of and extent to which vibrations of a particular frequency in one structure can be statistically related to those in the other structures (interstructural overlap); moreover, such values govern the extent to which vibrations within the same structure may overlap at nonnegligible values (intrastructural overlap).

After the frequency range is fixed, the sampling parameter L determines the total number of EVA descriptor elements; L should be maximized so as to reduce computational overhead and minimized to catch all the useful information. The optimal L value depends on the selected σ value.

Characteristic value of the Gaussian standard deviation σ is $10\,cm^{-1}$ (range 10–$20\,cm^{-1}$) and of the sampling increment L is $5\,cm^{-1}$ (range 2–$20\,cm^{-1}$), resulting in 800 descriptor variables.

The EVA descriptors are among $\rightarrow$ *3D descriptors*, independent of any molecular alignment, giving information about molecular size, shape, and electronic properties. Moreover, the EVA descriptors show only a moderate dependence on conformations.
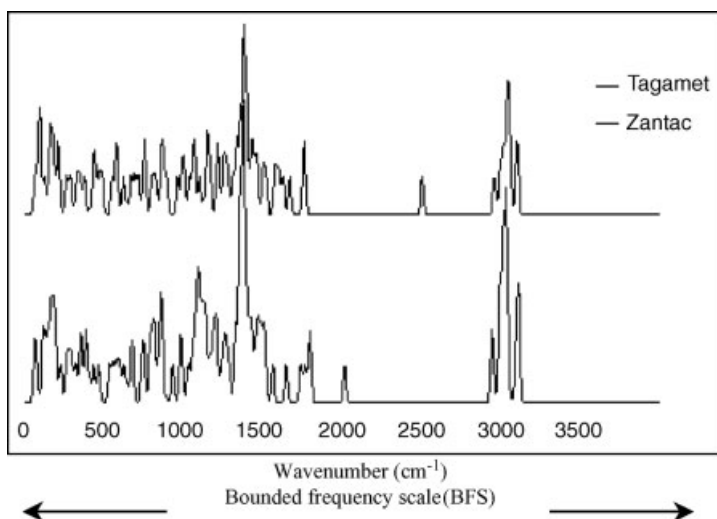

Figure E5 Bounded Frequency Scale with superimposed EVA descriptors for two compounds.

EEVA descriptors (or **Electronic EigenVAlue descriptors**) are $\rightarrow$ *vectorial descriptors* proposed as a modification of EVA [Tuppurainen, 1999a]. Semiempirical molecular orbital energies, that is, the eigenvalues of the Schrödinger equation, are used instead of the vibrational frequencies of the molecule. Each molecular orbital energy of the molecule is first projected onto a bounded energy scale (the range can be $-45 \div 10$ eV, but this depends on the quantum-chemical method used to calculate orbital energies). Then a Gaussian function of fixed standard deviation σ (0.50 eV was proposed) is centered at each MO energy projection resulting in a series identical and overlapping Gaussians. Once an appropriate sampling interval L (0.25 eV was proposed) has been chosen, the whole range is sampled; the EEVA descriptors at each sampling point x are defined as

$$\text{EEVA}_x = \sum_i \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-(x-\varepsilon_i)^2 / 2\sigma^2}$$

where the summation goes over all Gaussian functions and ε_i is the ith molecular orbital energy of the molecule. By using the parameter values defined above, this procedure provides a descriptor vector consisting of 220 (i.e., 55/0.25) elements so that dimensionality is much lower than that of the EVA descriptor vector.

 [Baumann, 1999; Benigni, Passerini *et al.*, 1999a, 1999b; Borosy, Balogh *et al.*, 2005; Devillers, 2000; Ford, Phillips *et al.*, 2004; Ginn, Turner *et al.*, 1997; Heritage, Ferguson *et al.*, 1998; Makhija and Kulkarni, 2001a; Takane and Mitchell, 2004; Tuppurainen and

Ruuskanen, 2000; Tuppurainen, Viisas *et al.*, 2002; Turner, Willett *et al.*, 1999; Turner and Willett, 2000a, 2000b]

➤ **EVA function** → EVA descriptors
➤ **evaluation set** ≡ *external evaluation set* → data set
➤ **Evans extended connectivity indices** → connectivity indices
➤ **EV$_{TYPE}$ descriptors** → van der Waals excluded volume method
➤ **EV$_{WHOLE}$ descriptors** → van der Waals excluded volume method
➤ **even/odd distance indices** → Wiener index
➤ **even/odd Wiener polynomial descriptors** → Wiener index
➤ **excess electron polarizability** → electric polarization descriptors
➤ **excess molar refractivity** → physico-chemical properties (⊙ molar refractivity)
➤ **Exner statistics** → regression parameters
➤ **expanded distance Cluj matrices** → expanded distance matrices
➤ **expanded distance indices** → expanded distance matrices

■ **expanded distance matrices** (≡ *distance-extended matrices*)
The original *expanded distance matrix*, denoted as Д, proposed by Tratch [Tratch, Stankevitch *et al.*, 1990] is a square symmetric $A \times A$ matrix representing a → *H-depleted molecular graph* with A vertices whose diagonal entries are equal to zero and each off-diagonal entry is defined as

$$[Д]_{ij} = \begin{cases} \mu_{ij} \cdot n_{ij} \cdot d_{ij} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

where d_{ij} denotes the → *topological distance* between vertices v_i and v_j, n_{ij} the number of external paths including the shortest path between the considered vertices with length $m \geq d_{ij}$, and μ_{ij} is the number of shortest paths between v_i and v_j that is equal to 1 for any pair of vertices in acyclic graphs. The number of external paths n_{ij} with respect to the shortest path i–j is just the same for each of the shortest paths connecting vertices v_i and v_j; therefore, the product $n_{ij} \times \mu_{ij}$ gives the total number of external paths in cyclic graphs.

Applying the → *Wiener operator Wi* to the expanded distance matrix, a molecular descriptor called **expanded Wiener number** $\tilde{W}$ is defined as

$$\tilde{W} \equiv Wi(Д) = \frac{1}{2} \cdot \sum_{i=1}^{A} \sum_{j=1}^{A} [Д]_{ij}$$

For acyclic graphs, the expanded distance matrix can be obtained simply as

$$Д = \mathbf{D} \otimes \mathbf{W}$$

where $\mathbf{D}$ and $\mathbf{W}$ denote the → *distance matrix* and the → *Wiener matrix*, respectively, and ⊗ indicates the → *Hadamard matrix product*. Therefore, the corresponding expanded Wiener number is calculated as

$$\tilde{W} \equiv \sum_{i=1}^{A-1} \sum_{j=i+1}^{A} d_{ij} \cdot N_i \cdot N_j$$

where N_i and N_j are the number of vertices on each side of the path i–j, including both vertices i and j, respectively. If only the bond contributions are considered ($d_{ij} = 1$), the expanded Wiener number coincides with the $\rightarrow$ *Wiener index*.

Example E6

Calculation of the expanded Wiener index and the Wiener index for 2-methylpentane.
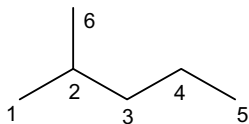
vertex pair (i,j)	d_{ij}	N_i	N_j	$d_{ij} \cdot N_i \cdot N_j$	vertex pair (i,j)	d_{ij}	N_i	N_j	$d_{ij} \cdot N_i \cdot N_j$
(1, 2)	1	1	5	5	(2, 6)	1	5	1	5
(1, 3)	2	1	3	6	(3, 4)	1	4	2	8
(1, 4)	3	1	2	6	(3, 5)	2	4	1	8
(1, 5)	4	1	1	4	(3, 6)	2	3	1	6
(1, 6)	2	1	1	2	(4, 5)	1	5	1	5
(2, 3)	1	3	3	9	(4, 6)	3	2	1	6
(2, 4)	2	3	2	12	(5, 6)	4	1	1	4
(2, 5)	3	3	1	9					

Expanded Wiener number:
$$\tilde{W} = 5 + 6 + 6 + 4 + 2 + 9 + 12 + 9 + 5 + 8 + 8 + 6 + 5 + 6 + 4 = 95$$

Wiener index: $W = d_{12} \cdot N_1 \cdot N_2 + d_{23} \cdot N_2 \cdot N_3 + d_{26} \cdot N_2 \cdot N_6 + d_{34} \cdot N_3 \cdot N_4 + d_{45} \cdot N_4 \cdot N_5 =$
$$= 5 + 9 + 5 + 8 + 5 = 32$$

Based on different powers of the topological distance, **generalized expanded Wiener numbers** $\tilde{W}^\lambda$, also called **Tratch–Stankevitch–Zefirov-type indices**, were proposed as [Klein and Gutman, 1999]

$$\tilde{W}^\lambda = \sum_{i=1}^{A-1} \sum_{j=i+1}^{A} d_{ij}^\lambda \cdot N_i \cdot N_j$$

where λ is any integer and the relation is valid only for trees.

Note that $\lambda = 0$ and $\lambda = 1$ result in, respectively, the $\rightarrow$ *hyper-Wiener index* and the expanded Wiener number; formally, for $\lambda \rightarrow -\infty$, the generalized expanded Wiener number should coincide with the well-known Wiener index. For cycle-containing graphs, the generalized expanded Wiener numbers were calculated as

$$\tilde{W}^\lambda = \sum_{i=1}^{A-1} \sum_{j=i+1}^{A} d_{ij}^\lambda \cdot \#_{ij}$$

where $\#_{ij}$ is the number of pairs of vertices (v_p, v_q) of the graph such that there is a geodesic (that is, the shortest path) between them containing both vertices v_i and v_j. The number $\#_{ij}$ is equal to the number of all shortest paths containing the path p_{ij} as a subpath, that is, the total number of external paths.

A generalization of the expanded distance matrix was proposed by Diudea [Diudea and Gutman, 1998] to define new matrices derived from the Hadamard matrix product between the distance matrix **D** and a general square $A \times A$ matrix **M** as

$$\mathbf{D\_M} = \mathbf{D} \otimes \mathbf{M}$$

If **M** is one among the $\rightarrow$ *Cluj matrices* then **expanded distance Cluj matrices** are obtained. Next, if **M** is the $\rightarrow$ *Szeged matrix* then the **expanded distance Szeged matrix** is derived; analogously, **expanded distance Szeged property matrices** [Minailiuc, Katona *et al.*, 1998] and **expanded distance walk matrices** are derived from $\rightarrow$ *Szeged property matrices* and $\rightarrow$ *walk matrices*.

From these matrices, two kinds of molecular indices are obtained. The **expanded distance indices**, denoted by DM_p or $D^U M_p$, depending on whether the matrix is symmetric or unsymmetric, are calculated by applying the $\rightarrow$ *Wiener operator Wi* to both expanded distance symmetric **M** and unsymmetric **UM** matrices:

$$DM_p \equiv Wi(\mathbf{D\_M}) = \frac{1}{2} \cdot \sum_{i=1}^{A} \sum_{j=1}^{A} [\mathbf{D\_M}]_{ij} \quad D^U M_p \equiv Wi(\mathbf{D\_UM}) \equiv \frac{1}{2} \cdot \sum_{i=1}^{A} \sum_{j=1}^{A} [\mathbf{D\_UM}]_{ij}$$
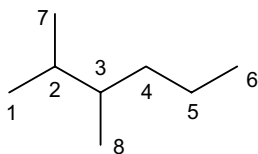
The **expanded square distance indices** are calculated, only from unsymmetrical **D_UM** matrices, by applying the $\rightarrow$ *orthogonal Wiener operator* $Wi^{\perp}$ as

$$D^2 M_p \equiv Wi^{\perp}(\mathbf{D\_UM}) = \sum_{i=1}^{A} \sum_{j=i}^{A} ([\mathbf{D\_UM}]_{ij} \cdot [\mathbf{D\_UM}]_{ji})$$

Note that $D^2 M_p$ indices involve square topological distances.

Example E7

Calculation of the expanded Wiener index and the Wiener index for 2,3-dimethylhexane.

D

	1	2	3	4	5	6	7	8	VS_i
1	0	1	2	3	4	5	2	3	20
2	1	0	1	2	3	4	1	2	14
3	2	1	0	1	2	3	2	1	12
4	3	2	1	0	1	2	3	2	14
5	4	3	2	1	0	1	4	3	18
6	5	4	3	2	1	0	5	4	24
7	2	1	2	3	4	5	0	3	20
8	3	2	1	2	3	4	3	0	18
CS_j	20	14	12	14	18	24	20	18	140

Wiener index $(W) = 70$

UCJD_p	1	2	3	4	5	6	7	8	VS_i
1	0	1	1	1	1	1	1	1	7
2	7	0	3	3	3	3	7	3	29
3	5	5	0	5	5	5	5	7	37
4	3	3	3	0	6	6	3	3	27
5	2	2	2	2	0	7	2	2	19
6	1	1	1	1	1	0	1	1	7
7	1	1	1	1	1	1	0	1	7
8	1	1	1	1	1	1	1	0	7
CS_j	20	14	12	14	18	24	20	18	140

Wiener index $(W) = 70$

hyper-Cluj-distance index $(CJD_p) = 143$

D_UCJD_p	1	2	3	4	5	6	7	8	VS_i
1	0	1	2	3	4	5	2	3	20
2	7	0	3	6	9	12	7	6	50
3	10	5	0	5	10	15	10	7	62
4	9	6	3	0	6	12	9	6	51
5	8	6	4	2	0	7	8	6	41
6	5	4	3	2	1	0	5	4	24
7	2	1	2	3	4	5	0	3	20
8	3	2	1	2	3	4	3	0	18
CS_j	44	25	18	23	37	60	44	35	286

$D^U CJD_p = 143$

$D^2 CJD_p = 605$

SCJD_p	1	2	3	4	5	6	7	8	VS_i
1	0	7	5	3	2	1	1	1	20
2	7	0	15	9	6	3	7	3	50
3	5	15	0	15	10	5	5	7	62
4	3	9	15	0	12	6	3	3	51
5	2	6	10	12	0	7	2	2	41
6	1	3	5	6	7	0	1	1	24
7	1	7	5	3	2	1	0	1	20
8	1	3	7	3	2	1	1	0	18
CS_j	20	50	62	51	41	24	20	18	286

hyper-Cluj-distance index $(CJD_p) = 143$

D_SCJD_p	1	2	3	4	5	6	7	8	VS_i
1	0	7	10	9	8	5	2	3	44
2	7	0	15	18	18	12	7	6	83
3	10	15	0	15	20	15	10	7	92
4	9	18	15	0	12	12	9	6	81
5	8	18	20	12	0	7	8	6	79
6	5	12	15	12	7	0	5	4	60
7	2	7	10	9	8	5	0	3	44
8	3	6	7	6	6	4	3	0	35
CS_j	44	83	92	81	79	60	44	35	518

$DCJD_p = 259$

Using the reciprocal distance matrix $\mathbf{D}^{-1}$ instead of the simple distance matrix in the Hadamard matrix product, **expanded reciprocal distance matrices H_M** are obtained as

$$\mathbf{H\_M} = \mathbf{D}^{-1} \otimes \mathbf{M}$$

where $\mathbf{M}$ can be any square $A \times A$ matrix as defined above.

From these matrices, the **expanded reciprocal distance indices** (HM_p and $H^U M_p$) and **expanded reciprocal square distance indices** ($H^2 M_p$) are derived as

$$HM_p \equiv Wi(\mathbf{H\_M}) = \frac{1}{2} \cdot \sum_{i=1}^{A} \sum_{j=1}^{A} [\mathbf{H\_M}]_{ij} \qquad H^U M_p \equiv Wi(\mathbf{H\_UM}) = \frac{1}{2} \cdot \sum_{i=1}^{A} \sum_{j=1}^{A} [\mathbf{H\_UM}]_{ij}$$

and, only for unsymmetrical $\mathbf{H\_UM}$ matrices,

$$H^2 M_p \equiv Wi^{\perp}(\mathbf{H\_UM}) = \sum_{i=1}^{A} \sum_{j=i}^{A} ([\mathbf{H\_UM}]_{ij} \cdot [\mathbf{H\_UM}]_{ji})$$

Moreover, following the same procedure, other expanded matrices are defined [Minailiuc, Katona *et al.*, 1998] replacing the topological distance matrix $\mathbf{D}$ by the → *geometry matrix* $\mathbf{G}$ as

$$\mathbf{G\_M} = \mathbf{G} \otimes \mathbf{M} \quad \text{and} \quad \mathbf{G^{-1}\_M} = \mathbf{G^{-1}} \otimes \mathbf{M}$$

where $\mathbf{M}$ can be any square $A \times A$ matrix as defined above and $\mathbf{G^{-1}}$ is the → *reciprocal geometry matrix*. Therefore, $\mathbf{G\_M}$ and $\mathbf{G^{-1}\_M}$ matrices can be called **expanded geometric distance matrices** and **expanded reciprocal geometric distance matrices**, respectively. The corresponding molecular indices are defined as

$$GM_p \equiv Wi(\mathbf{G\_M}) = \frac{1}{2} \cdot \sum_{i=1}^{A} \sum_{j=1}^{A} [\mathbf{G\_M}]_{ij} \qquad G^U M_p \equiv Wi(\mathbf{G\_UM}) = \frac{1}{2} \cdot \sum_{i=1}^{A} \sum_{j=1}^{A} [\mathbf{G\_UM}]_{ij}$$

$$H_G M_p \equiv Wi(\mathbf{G^{-1}\_M}) = \frac{1}{2} \cdot \sum_{i=1}^{A} \sum_{j=1}^{A} [\mathbf{G^{-1}\_M}]_{ij}$$

$$H_G^U M_p \equiv Wi(\mathbf{G^{-1}\_UM}) = \frac{1}{2} \cdot \sum_{i=1}^{A} \sum_{j=1}^{A} [\mathbf{G^{-1}\_UM}]_{ij}$$

and, only for unsymmetrical $\mathbf{G\_UM}$ and $\mathbf{G^{-1}\_UM}$ matrices,

$$G^2 M_p \equiv Wi^{\perp}(\mathbf{G\_UM}) = \sum_{i=1}^{A} \sum_{j=i}^{A} ([\mathbf{G\_UM}]_{ij} \cdot [\mathbf{G\_UM}]_{ji})$$

$$H_G^2 M_p \equiv Wi^{\perp}(\mathbf{G^{-1}\_UM}) = \sum_{i=1}^{A} \sum_{j=i}^{A} ([\mathbf{G^{-1}\_UM}]_{ij} \cdot [\mathbf{G^{-1}\_UM}]_{ji})$$

The symbols of the molecular descriptors derived from the most common expanded distance matrices are listed in the Table E12.

Table E12 Wiener-type indices derived from some expanded distance matrices.

Molecular descriptor	. . . obtained from the matrix . . .
$\tilde{W}$	Expanded distance matrix
HW_p	Expanded reciprocal distance path-Wiener matrix
GW_p	Expanded geometric distance path-Wiener matrix
$H_G W_p$	Expanded reciprocal geometric distance path-Wiener matrix
$DCJD_p$	Expanded distance symmetric path-Cluj-distance matrix
$D^U CJD_p$ and $D^2 CJD_p$	Expanded distance unsymmetric path-Cluj-distance matrix
$HCJD_p$	Expanded reciprocal distance symmetric path-Cluj-distance matrix
$H^U CJD_p$ and $H^2 CJD_p$	Expanded reciprocal distance unsymmetric path-Cluj-distance matrix
$GCJD_p$	Expanded geometric distance symmetric path-Cluj-distance matrix
$G^U CJD_p$ and $G^2 CJD_p$	Expanded geometric distance unsymmetric path-Cluj-distance matrix
$H_G CJD_p$	Expanded reciprocal geometric distance symmetric path-Cluj-distance matrix
$H_G^U CJD_p$ and $H_G^2 CJD_p$	Expanded reciprocal geometric distance unsymmetric path-Cluj-distance matrix
$DCJ\Delta_p$	Expanded distance symmetric path-Cluj-detour matrix
$D^U CJ\Delta_p$ and $D^2 CJ\Delta_p$	Expanded distance unsymmetrical path-Cluj-detour matrix
$HCJ\Delta_p$	Expanded reciprocal distance symmetric path-Cluj-detour matrix
$H^U CJ\Delta_p$ and $H^2 CJ\Delta_p$	Expanded reciprocal distance unsymmetrical path-Cluj-detour matrix
$GCJ\Delta_p$	Expanded geometric distance symmetric path-Cluj-detour matrix
$G^U CJ\Delta_p$ and $G^2 CJ\Delta_p$	Expanded geometric distance unsymmetrical path-Cluj-detour matrix

(Continued)

Table E12 (*Continued*)

Molecular descriptor	. . . obtained from the matrix . . .
$H_G CJ\Delta_p$	Expanded reciprocal geometric distance symmetric path-Cluj-detour matrix
$H_G^U CJ\Delta_p$ and $H_G^2 CJ\Delta_p$	Expanded reciprocal geometric distance unsymmetrical path-Cluj-detour matrix
DSZ_p	Expanded distance symmetric path-Szeged matrix
$D^U SZ_p$ and $D^2 SZ_p$	Expanded distance unsymmetrical path-Szeged matrix
HSZ_p	Expanded reciprocal distance symmetric path-Szeged matrix
$H^U SZ_p$ and $H^2 SZ_p$	Expanded reciprocal distance unsymmetrical path-Szeged matrix
GSZ_p	Expanded geometric distance symmetric path-Szeged matrix
$G^U SZ_p$ and $G^2 SZ_p$	Expanded geometric distance unsymmetrical path-Szeged matrix
$H_G SZ_p$	Expanded reciprocal geometric distance symmetric path-Szeged matrix
$H_G^U SZ_p$ and $H_G^2 SZ_p$	Expanded reciprocal geometric distance unsymmetrical path-Szeged matrix

If the calculation of indices in Table E12 is performed by summing only the matrix entries corresponding to pairs of adjacent vertices, then similar edge-defined indices can be calculated.

📖 [Diudea, Pârv *et al.*, 1997b]

➤ **expanded distance Szeged matrix** → expanded distance matrices
➤ **expanded distance Szeged property matrices** → expanded distance matrices
➤ **expanded distance walk matrices** → expanded distance matrices
➤ **expanded geometric distance matrices** → expanded distance matrices
➤ **expanded matrices** → matrices of molecules
➤ **expanded reciprocal distance indices** → expanded distance matrices
➤ **expanded reciprocal distance matrices** → expanded distance matrices
➤ **expanded reciprocal geometric distance matrices** → expanded distance matrices
➤ **expanded reciprocal square distance indices** → expanded distance matrices
➤ **expanded square distance indices** → expanded distance matrices
➤ **expanded Wiener number** → expanded distance matrices
➤ **expected square error** ≡ *mean square error* → regression parameters
➤ **experimental design** → chemometrics

◼ **experimental measurements**
Experimental measurements are the basis from which numerical or graphical information can be extracted by experiments. An experiment is a well-defined operational procedure that measures a quantity for a given sample; hence, experimental quantities are quantities measured by experimental measurement.

→ *Physico-chemical properties* and spectroscopic signals constitute the most important class of experimental chemical measurements, also playing a fundamental role as → *molecular descriptors* both for their availability as well as their interpretability. Examples of physico-chemical measurable quantities are refractive indices, molar refractivities, parachors, densities, solubilities, partition coefficients, dipole moments, chemical shifts, retention times, spectroscopic signals, rate constants, equilibrium constants, vapour pressures, boiling and melting points, acid dissociation constants, and so on [Baum, 1997; Horvath, 1992; Lyman, Reehl *et al.*, 1982; Reid, Prausnitz *et al.*, 1988].

$\rightarrow$ *Biological activities*, $\rightarrow$ *toxicological indices*, $\rightarrow$ *environmental indices* are other basic experimental quantities that are often considered as responses in QSAR modeling. Examples of biological measurable quantities are effects due to a concentration of a compound, binding affinities, toxicities, inhibition constants, carcinogenicity, mutagenicity, and teratogenicity. Examples of environmental measurable quantities are biochemical oxygen demand, chemical oxygen demand, biodegradability, bioconcentration factors, atmospheric residence time, volatilization from soil, and rate constants of atmospheric degradation reactions.

From a theoretical point of view, experimental properties $\mathcal{P}$ can be distinguished with respect to their behavior in a system S. A molecular property $\mathcal{P}$ may be categorized in terms of its behavior under the hypothesis that a system $S = \mathcal{A} \cup \mathcal{B}$ breaks up into two separate non-interacting subsystems $\mathcal{A}$ and $\mathcal{B}$ [Trinajstić, Randić *et al.*, 1986].

The physical behavior of the property has at least four mathematical possibilities of interest:

(1) $\quad \mathcal{P}(S) = \mathcal{P}(\mathcal{A}) + \mathcal{P}(\mathcal{B})$
(2) $\quad \mathcal{P}(S) = \mathcal{P}(\mathcal{A}) \quad$ or $\quad \mathcal{P}(\mathcal{B})$
(3) $\quad \mathcal{P}(S) = \mathcal{P}(\mathcal{A}) \cdot \mathcal{P}(\mathcal{B})$
(4) $\quad \mathcal{P}(S) = \partial\mathcal{P}(\mathcal{A}) \cdot \mathcal{P}(\mathcal{B}) - \mathcal{P}(\mathcal{A}) \cdot \partial\mathcal{P}(\mathcal{B})$

where $\mathcal{A} \cap \mathcal{B} = \varnothing$. These properties are termed (1) *additive*, (2) *constantive*, (3) *multiplicative*, and (4) *derivative*, respectively. Additive and constantive properties correspond to those properties called, in physical language, extensive and intensive properties.

Many physico-chemical properties and biological activities seem to fall within the domain of additive properties. Examples of constantive properties are local molecular properties, such as dissociation energy for a localized bond or ionization potential. Characteristic multiplicative "properties" are wave functions, Kekulé structure counts, and probabilities. The derivative properties are associated with the corresponding multiplicative property $\mathcal{P}$.

The choice of suitable QSAR/QSPR approaches as well as effective molecular descriptors depends on the characteristic behavior of the property studied.

For each experimental quantity, several estimation methods are known; moreover, for many of them, theoretical methods and/or empirical models were proposed to obtain reliable estimates avoiding experimental measurements.

In any case, experimental values or their estimated values are commonly used as molecular descriptors (i.e., as predictors in X block) or constitute the response (i.e., in Y block) that has to be modeled by other descriptors, that is, reproduced by theoretical models.

Some experimental quantities, such as spectra, need to be transformed in some way before they are used as descriptors. For example, infrared spectra signals (IR spectra) sampled at $10\,cm^{-1}$ in the fingerprint region (1500–600 cm^{-1}) were used as $\rightarrow$ *vectorial descriptors*, each spectrum being scaled in the range 0–100 [Benigni, Passerini *et al.*, 1999a].

📖 [Chaumat, Chamel *et al.*, 1992; Crebelli, Andreoli *et al.*, 1992; Dearden, 1990; Gasteiger, 1988; Horvath, 1988, 1992; Jochum, Hicks *et al.*, 1988; Müller and Klein, 1991]

➢ **explanatory variables** $\equiv$ *independent variables* $\rightarrow$ data set
➢ **exploratory data analysis** $\rightarrow$ chemometrics
➢ **exponential cost function** $\rightarrow$ regression parameters

> ➢ **exponential product-sum connectivities** → exponential sum connectivities
> ➢ **exponential similarity index** → quantum similarity

■ **exponential sum connectivities**

These are → *local vertex invariants* proposed by Balaban [Balaban and Catana, 1993] with the aim of obtaining high discrimination among the vertices of a → *H-depleted molecular graph*. They are denoted by c_i and are expressed in logarithmic units as

$$\log c_i = \left(\sum_{k=1}^{\eta_i} k^z \cdot \prod_{v_j \in V_{ik}} G_{jk} \right) \cdot \log G_i$$

where the sum runs over all the distances from vertex v_i to any other vertex in the graph up to its → *atom eccentricity* η_i, that is, the maximum distance, and it involves the products of G values of the atoms v_j, located at distance k from the vertex v_i and constituting the vertex subset V_{ik}; G_i is the value of local invariant G for the considered vertex v_i defined as

$$\log G_i = \left(\prod_{j=1}^{\delta_i} g_j \right) \cdot \log g_i \quad \text{being} \quad g_i = (\delta_i)^{-1/2} \cdot \left(\sum_{j=1}^{\delta_i} \delta_j \right)^{-1}$$

where the product involves all g values of the δ_i first neighbors of the vertex v_i, the local invariant g being defined as a function of the → *vertex degree* δ_i of the considered atom and the vertex degrees δ_j of its first neighbors. The exponent z in the formula of the c invariant can be either equal to $+1$, leading to **distance-enhanced exponential sum connectivities** or equal to -1 leading to **distance-normalized exponential sum connectivities**.

Analogously, other highly discriminating local invariants were proposed, denoted by c'_i and called **exponential product-sum connectivities**:

$$\log c'_i = \left(\sum_{k=1}^{\eta_i} k^z \cdot \prod_{v_j \in V_{ik}} G'_{jk} \right) \cdot \log G'_i$$

where

$$\log G'_i = \left(\prod_{j=1}^{\delta_i} g'_j \right) \cdot \log g'_i \quad \text{and} \quad g'_i = (\delta_i)^{-1/2} \cdot \left(\sum_{j=1}^{\delta_i} \delta_j + \prod_{j=1}^{\delta_i} \delta_j \right)^{-1}$$

Exponential sum connectivities c_i and c'_i take real values in the 0–1 range. Larger values are assigned to vertices that have higher degrees, are closer to the → *graph center*, and are closer to a vertex of higher degree.

By summing the exponential sum connectivities of all vertices, the corresponding molecular descriptors are derived:

$$XC = \sum_{i=1}^{A} c_i \quad \text{and} \quad XC' = \sum_{i=1}^{A} c'_i$$

where A is the number of vertices in the molecular graph. These topological indices decrease asymptotically toward zero with increasing number of vertices in linear chain graphs, while they increase toward infinity with increasing the number of vertices in highly branched graphs.

Exponential sum connectivities c_i and c'_i can also be calculated for graph fragments or organic substituents X, considering the dimer molecule X–X (e.g., for the ethyl group *n*-butane is taken)

[Balaban and Catana, 1994]. The values of c_i or c'_i are calculated for each half of such a dimer and assigned as LOVIs to each fragment. Moreover, the following descriptors for the whole fragment were proposed:

$$G_1 = \sum_i e^{-d_i} \cdot c_i \qquad\qquad G'_1 = \sum_i e^{-d_i} \cdot c'_i$$
$$G_2 = \sum_i 2^{-d_i} \cdot c_i \qquad\qquad G'_2 = \sum_i 2^{-d_i} \cdot c'_i$$
$$G_3 = \sum_i 10^{1-d_i} \cdot c_i \qquad\qquad G'_3 = \sum_i 10^{1-d_i} \cdot c'_i$$
$$G_4 = \sum_i d_i^{-3} \cdot c_i \qquad\qquad G'_4 = \sum_i d_i^{-3} \cdot c'_i$$

where the summation runs over all the vertices of the fragment, d is the topological distance of the considered atom from the root vertex.

For groups containing heteroatoms or multiple bonds, the LOVIs of each vertex are multiplied by the parameter f_i:

$$f_i = \frac{R_i}{R_{C_{sp^3}}}$$

where R_i is the covalent radius of the ith atom in its hybridization state and $R_{C_{sp^3}}$ is the covalent radius of C_{sp^3} (0.77 Å).

Moreover, for monocyclic substituents, the following descriptor was proposed:

$$G_1 = \sum_i c_i \cdot f_i \cdot \exp(-d_i - d_i/N_{BR})$$

where N_{BR} is the number of ring adjacencies (i.e., six of cyclohexane, nine for benzene). An analogous formula holds for G'_1 on replacing c_i by c'_i.

➢ **extended adjacency ID number** → ID numbers

■ **extended adjacency matrices**

The extended adjacency matrices **EA** are → *weighted adjacency matrices $A \times A$* whose elements are defined as a function of → *local vertex invariants* of the → *adjacency matrix* **A** and of some → *atomic properties* [Yang, Xu *et al.*, 1994]. The defined functions aim at removing degeneracy of the entries of the adjacency matrix that is a binary matrix and resemble to some extent the function used in defining the → *χ matrix*.

The **extended vertex adjacency matrix** (or simply **extended adjacency matrix**) is an adjacency matrix **EA** whose entries are defined as

$$[\mathbf{EA}]_{ij} = \begin{cases} a_{ij} \cdot \dfrac{\delta_i/\delta_j + \delta_j/\delta_i}{2} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

where a_{ij} are the entries of the adjacency matrix and δ is the → *vertex degree*.

A correction factor can be introduced to account for heteroatoms, such as the → *atomic electronegativity* χ_i; then, the entries of the **heteroatom-corrected extended adjacency matrix EA^h** are the following:

$$[\mathbf{EA}^h]_{ij} = \begin{cases} a_{ij} \cdot \dfrac{\delta_i^h/\delta_j^h + \delta_j^h/\delta_i^h}{2} & \text{if } i \neq j \\ \chi_i & \text{if } i = j \end{cases}$$

where $\delta_i^h = \delta_i \cdot \chi_i$.

A further extension of the extended adjacency matrix can also be made to consider → *bond multiplicity*; the entries of the **heteroatom/multiplicity-corrected extended adjacency matrix**, denoted by $\mathbf{EA}^{hb}$, are the following:

$$
[\mathbf{EA}^{hb}]_{ij} =
\begin{cases}
\dfrac{1}{2} \cdot \left(\dfrac{\delta_i^h + 1 - 1/\pi_{ij}^*}{\delta_j^h + 1 - 1/\pi_{ij}^*} + \dfrac{\delta_j^h + 1 - 1/\pi_{ij}^*}{\delta_i^h + 1 - 1/\pi_{ij}^*} \right) & \text{if} \quad (i,j) \in \mathcal{E}(G) \\
\chi_i & \text{if} \quad i = j \\
0 & \text{if} \quad (i,j) \notin \mathcal{E}(G)
\end{cases}
$$

where the increment to the modified vertex degree being due to the multiplicity is

$$
\begin{cases}
0.00 & \text{if} \quad \pi^* = 1 \\
0.33 & \text{if} \quad \pi^* = 1.5 \\
0.50 & \text{if} \quad \pi^* = 2 \\
0.67 & \text{if} \quad \pi^* = 3
\end{cases}
$$

where π_{ij}^* is the → *conventional bond order*.

Note. In the original paper, the heteroatom/multiplicity degrees are defined as the following:

$$
\begin{cases}
\delta_i^{hb} = \delta_i^h & \text{if} \quad \pi^* = 1 \\
\delta_i^{hb} = \delta_i^h + 1/2 & \text{if} \quad \pi^* = 2 \\
\delta_i^{hb} = \delta_i^h + 1/3 & \text{if} \quad \pi^* = 3
\end{cases}
$$

However, by this definition, the ranking of the corrected degrees due to different multiplicities is doubtful, the triple bond increment being intermediate to single and double bonds.

From the extended adjacency matrix defined above two → *spectral indices*, called → *extended adjacency matrix indices*, were proposed as molecular descriptors. Moreover, the → *extended adjacency ID number* was derived from an extended adjacency matrix defined in terms of atomic covalent radii and local vertex invariants computed from → *layer matrices*.

The **extended edge adjacency matrix**, denoted as $\mathbf{E^E A}$, was also defined by analogy with the extended vertex adjacency matrix. This is a symmetric $B \times B$ matrix, B being the number of edges in the graph, where the vertex degrees are replaced by the → *edge degrees* ε [Janežič, Miličević *et al.*, 2007]. The extended edge adjacency matrix of a molecular graph G is the extended vertex adjacency matrix of the corresponding → *line graph* L(G).

➢ **extended adjacency matrix indices** → spectral indices
➢ **extended adjacency matrix** ≡ *extended vertex adjacency matrix* → extended adjacency matrices
➢ **Extended Connection Table Representation** → molecular graph
➢ **extended connectivity** → canonical numbering (⊙ Morgan's extended connectivity algorithm)
➢ **extended connectivity algorithm** ≡ *Morgan's extended connectivity algorithm* → canonical numbering
➢ **Extended Connectivity FingerPrints** → substructure descriptors (⊙ fingerprints)
➢ **extended connectivity indices** → canonical numbering (⊙ Morgan's extended connectivity algorithm)

- ➢ **extended edge adjacency matrix** → extended adjacency matrices
- ➢ **extended edge connectivity indices** → edge adjacency matrix
- ➢ **extended Ivanciuc–Balaban operator** → Balaban distance connectivity index
- ➢ **extended local information on distances** → topological information indices
- ➢ **extended Madan degree** → vertex degree
- ➢ **Extended Set of Smallest Rings** → ring descriptors
- ➢ **extended vertex adjacency matrix** → extended adjacency matrices
- ➢ **extended vertex degree** ≡ *extended connectivity* → canonical numbering (⊙ Morgan's extended connectivity algorithm)
- ➢ **extended Wiener–Hosoya indices** → Wiener index
- ➢ **external evaluation set** → data set
- ➢ **External Factor Variable Connectivity Indices** → variable descriptors
- ➢ **external fragment topological indices** → fragment topological indices
- ➢ **external validation** → validation techniques

◼ **extrathermodynamic approach**

Sometimes used as a synonym of → *Hansch analysis*, the extrathermodynamic approach refers to models based on empirical relationships of → *physico-chemical properties* with thermodynamic parameters such as free energies, enthalpies, and entropies for various reactions.

These relationships, based on thermodynamic parameters but not requiring the formal thermodynamic theory, are therefore "extrathermodynamic."

The Hammett, Taft, and Hansch–Fujita equations defining electronic, steric, and hydrophobic constants are examples of extrathermodynamic relationships, being based on logarithms of rate or equilibrium constants, that is, free energy-related quantities of standard organic reactions of congeneric compound series. Since these correlation equations are often linear with respect to at least one variable, they are called **Linear Free Energy Relationships** (LFER) [Fujita, 1990; Mekenyan and Bonchev, 1986; Wells, 1968b].

Between 1961–1964 Hansch and coworkers used, for the first time, an extrathermodynamic approach to mathematically relate biological activity to the physico-chemical properties of molecules.

The basic assumption is that the introduction of different substituents into a reference compound modifies its biological activity, which can be expressed to a first-order approximation by the following relationship:

$$\Delta K = f(\Delta \Phi_1, \Delta \Phi_2, \dots, \Delta \Phi_J)$$

where Φ are molecular physico-chemical properties, usually electronic, steric, and hydrophobic properties, and K (or k) is the equilibrium (or rate) constant of the biological interaction.

📖 [Abraham, Whiting *et al.*, 1998; Chapman and Shorter, 1978; Charton, 1978a; Drmanić, Jovanović *et al.*, 2000; Gironés and Carbó-Dorca, 2003; Hansch, Quinlan *et al.*, 1968; Jenkins, Samuel *et al.*, 1995; Konovalov, Coomans *et al.*, 2007; Kubinyi, 1993b; Ohlenbusch and Frimmel, 2001; Platts, Abraham *et al.*, 2000; Platts, Butina *et al.*, 1999; Ponec, Gironés *et al.*, 2002; Roberts, 1995; Simón-Manso, 2005; Verloop, 1972; Zissimos, Abraham *et al.*, 2002c]

- ➢ *E weighting scheme* → weighting schemes