

H

- **HACA index** → charged partial surface area descriptors
- **Hadamard matrix product** → algebraic operators
- **hafnian** → algebraic operators (⊙ determinant)
- **half-life time** → environmental indices
- **HLOGP** → lipophilicity descriptors
- **Hamann similarity coefficient** → similarity/diversity (⊙ Table S9)
- **Hamiltonian circuit** → graph
- **Hamiltonian path** → graph
- **Hammett electronic constant** → Hammett equation

■ Hammett equation

Proposed by Hammett in 1937 [Hammett, 1937, 1938; Johnson, 1973], the Hammett equation is defined for the rate constants k and equilibrium constants K of reactions of *meta*- and *para*-substituted benzoic acid derivatives:

$$\rho \cdot \sigma = \log \left(\frac{k_X}{k_H} \right) \quad \text{and} \quad \rho \cdot \sigma = \log \left(\frac{K_X}{K_H} \right)$$

where the constants k_H and K_H refer to an unsubstituted compound (i.e., with hydrogen in the substitution site), while k_X and K_X refer to a *meta*- or *para*-X-substituted compound. *Ortho*-substituents are less used as the electronic effect could be complicated by steric interactions.

The substituent constant σ , called **Hammett electronic constant**, depends only on the nature and position of the substituent and is related to the electronic effect the substituent has on the rate or equilibrium of the reaction, relative to the unsubstituted compound.

The **reaction constant** ρ depends upon the reaction, the conditions under which it is studied and the nature of the reaction series. The magnitude of ρ gives the susceptibility of a given reaction to polar substituents. Large positive values are obtained from all base-catalyzed reactions, while for acid-catalyzed reactions ρ values are of variable sign but in all cases quite small.

Therefore, σ is an electronic descriptor of the substituent estimated by measured rate or equilibrium constants of a reaction, under the control parameter ρ .

From the Hammett equation, several $\sigma \rightarrow$ *electronic substituent constants* are derived from different reactions and different experimental conditions; a modification of the Hammett equation was defined as the \rightarrow *Yukawa–Tsumo equation*.

☞ [Hammett, 1935; Jaffé, 1953; Yamamoto and Otsu, 1967; Shorter, 1978; Roberts, 1995; Suresh and Gadre, 1998; Popelier, 1999; Drmanić, Jovanović *et al.*, 2000; Lin, Yin *et al.*, 2003; Verma, Kapur *et al.*, 2003; Liu, Fu *et al.*, 2004; Simón-Manso, 2005; Smith and Popelier, 2005]

- **Hammett substituent constants** \equiv *electronic substituent constants*
- **Hamming distance** \rightarrow similarity/diversity (☉ Table S10)
- **Hamming similarity coefficient** \rightarrow similarity/diversity (☉ Table S9)
- **Hancock steric constant** \equiv *corrected Taft steric constant* \rightarrow steric descriptors (☉ Taft steric constant)
- **Hannan–Quinn ϕ -criterion** \rightarrow regression parameters (☉ Table R1)

■ Hansch analysis

Derived from physical organic chemistry and the \rightarrow *Hammett equation*, it can be considered the first approach to modern QSAR studies. Proposed by Hansch and coworkers in the early 1960s [Hansch, Maloney *et al.*, 1962; Hansch, Muir *et al.*, 1963; Hansch and Fujita, 1964; Hansch, Deutsch *et al.*, 1965; Hansch and Anderson, 1967; Hansch, 1969, 1971, 1978], it is the investigation of the quantitative relationships between the biological activity of a series of compounds and their physico-chemical parameters representing hydrophobic, electronic, steric, and other effects using multivariate regression methods [Kubinyi, 1993b].

Hansch analysis assumes that variations in the magnitude of a certain biological activity exhibited by a series of bioactive compounds can be correlated to variations in different physico-chemical factors associated with their structure. Therefore, the basic QSAR equation in the Hansch analysis is defined as

$$\text{biological activity} = f(\Phi_1, \Phi_2, \dots, \Phi_J)$$

where Φ are \rightarrow *physico-chemical properties* of congeneric compounds having a common skeleton but varying substituents. Together with the most significant parameters, factors for hydrogen bonding, van der Waals and charge-transfer forces, etc. can be used, depending on the situation.

The biological activity is usually defined as $\log(1/C)$, where C is the molar concentration of a compound producing a fixed effect.

Hansch analysis tries to correlate biological activity with physico-chemical properties by linear and nonlinear regression analysis, finding property-activity relationship models. A **Craig plot** is a plot of two substituent parameters (e.g., Hansch–Fujita π and Hammett σ values).

The simplest Hansch analysis is based on the **Hansch linear model** [Kubinyi, 1988b], defined as

$$\hat{y}_i = b_0 + \sum_{j=1}^J b_j \cdot \Phi_{ij}$$

where Φ_{ij} represents the j th physico-chemical property of the i th compound, b_j the regression coefficients, and J the total number of considered properties. The intercept b_0 corresponds to a theoretical biological activity of a compound whose all the property values are zero. This condition is approximately fulfilled for a hydrogen substituted reference compound as several property values are normalized to zero. Depending on the regression coefficient significance, some factors can result not relevant.

For example, a typical Hansch linear equation for monosubstituted derivatives is as the following:

$$\log(1/C) = b_0 + b_1 \cdot \pi + b_2 \cdot \sigma + b_3 \cdot E_s$$

where π indicates \rightarrow *Hansch–Fujita hydrophobic substituent constants*, σ the \rightarrow *electronic substituent constants*, and E_s the \rightarrow *Taft steric constant*.

The j th molecular property of the i th compound Φ_{ij} can be defined as the sum of the values of the substituent constant ϕ of type j over all substituents of that compound:

$$\Phi_{ij} = \sum_{s=1}^S \sum_{k=1}^{N_s} \phi_{ks,j} \cdot I_{i,ks}$$

where $I_{i,ks}$ are \rightarrow *indicator variables* (such as in \rightarrow *Free–Wilson analysis*) indicating the presence, that is, $I_{i,ks} = 1$, and absence, that is, $I_{i,ks} = 0$, of the k th substituent in the s th site for the i th compound; $\phi_{ks,j}$ the j th property of the k th substituent in the s th site; S the number of substitution sites; and N_s the number of group substituents in the s th site. As in each site only one substituent is present for a given compound, S is the total number of contributions to the considered molecular property. Alternatively, the j th molecular property Φ_{ij} of the i th compound is defined as the j th global molecular property such as \rightarrow *log P* or some global \rightarrow *steric descriptors*.

Both the molecular physico-chemical properties Φ obtained by the above relationship and the \rightarrow *substituent constants* ϕ are usually known as **Hansch descriptors** [Hansch, Leo *et al.*, 1995].

Also to take into account some nonlinear contributions of the properties [Hansch and Clayton, 1973; Kubinyi, 1993b], a **Hansch parabolic model** is defined as

$$\hat{y}_i = b_0 + \sum_{j=1}^J b_j \cdot \Phi_{ij} + \sum_{j=1}^J b''_j \cdot \Phi_{ij}^2$$

Among the quadratic terms, usually the most used is $(\log P)^2$, to mimic the nonlinear behavior of the interchange between a two-phase system (e.g., aqueous/organic system), that is, too low or too high lipophilicity values act as limiting factor. The most common parabolic model is specifically defined as

$$\log(1/C) = b_0 - b_1 \cdot (\log P)^2 + b_2 \cdot \log P + b_3 \cdot \sigma$$

The general **Hansch nonlinear model** is defined as

$$\hat{y}_i = b_0 + \sum_{j=1}^J b_j \cdot \Phi_{ij} + \sum_{j=1}^J \sum_{j'=j}^J b''_{jj'} \cdot \Phi_{ij} \cdot \Phi_{ij'}$$

also taking into account the combined effects of the properties, even if not usually considered.

Besides the nonlinear models and, specifically, the parabolic model, other models were proposed for nonlinear dependence of the biological response on hydrophobic interactions. Among them, the most important are the **Hansch bilinear models** [Kubinyi, 1977, 1979] such as

$$\log(1/C) = b_0 + b_1 \cdot \log P - b_2 \cdot \log(\beta \cdot P + 1)$$

Special cases of such bilinear models are the **McFarland model** [McFarland, 1970], derived for $b_2 = 2 \times b_1$ and $\beta = 1$ and the **Higuchi–Davis model** [Higuchi and Davis, 1970], for $b_2 = 1$ and

$\beta = V_{\text{lip}}/V_{\text{aq}}$, which is the ratio of the volume of the lipid phase V_{lip} over the volume of the aqueous phase V_{aq} .

The Hansch linear model is related to the \rightarrow *Fujita–Ban model* when, in both models, the hydrogen substituted compound is taken as the reference compound; each Fujita–Ban regression coefficient b_{ks} corresponds to the Hansch equation for a single substituent:

$$b_{ks} \approx \sum_{j=1}^J b_j \cdot \phi_{ks,j}$$

where J is the number of considered properties (e.g., lipophilic, electronic, and steric properties) and $\phi_{ks,j}$ is the j th substituent group property for the k th substituent in the s th site. This relationship means that the group contribution b_{ks} in the Fujita–Ban model of the k th substituent in the s th site is numerically equivalent to the weighted sum of all the physico-chemical properties of that substituent [Kubinyi and Kehrnhahn, 1976]. Substituting the previous relationship in the Fujita–Ban model, it can be observed that the two models are closely related:

$$\begin{aligned} \hat{y}_i &= b_0 + \sum_{s=1}^S \sum_{k=1}^{N_s} b_{ks} \cdot I_{i,ks} = b_0 + \sum_{s=1}^S \sum_{k=1}^{N_s} \sum_{j=1}^J b_j \cdot \phi_{ks,j} \cdot I_{i,ks} \\ &= b_0 + \sum_{j=1}^J b_j \cdot \sum_{s=1}^S \sum_{k=1}^{N_s} \phi_{ks,j} \cdot I_{i,ks} = b_0 + \sum_{j=1}^J b_j \cdot \Phi_{ij} \end{aligned}$$

In particular, the Fujita–Ban group contributions implicitly contain all the possible physico-chemical contributions of a substituent; as a consequence, the Fujita–Ban models always give an upper limit of correlation, which can be achieved by Hansch linear models.

Hansch–Free–Wilson mixed models were also proposed [Kubinyi, 1976a] by combining the two approaches in a single model. A quadratic term accounting for hydrophobic interactions (usually $\log P$ or π Hansch–Fujita constant) can be added to the Free–Wilson (or Fujita–Ban) model as

$$\hat{y}_i = b_0 + \sum_{s=1}^S \sum_{k=1}^{N_s} b_{ks} \cdot I_{i,ks} + b' \cdot (\log P_i)^2$$

where the first part is the Free–Wilson model, S and N_s , respectively, the number of substitution sites and substituent groups in each s th site, and $I_{i,ks}$ is an indicator variable for the i th compound denoting presence (1) or absence (0) of the k th group in the s th site.

Mixed models can also be obtained by mixing the two approaches, each describing a different group of substituents:

$$\hat{y}_i = b_0 + \sum_{s=1}^S \sum_{k=1}^{N_s} b_{ks} \cdot I_{i,ks} + \sum_{j=1}^J b'_j \cdot \Phi_{ij} + b'' \cdot (\log P_i)^2$$

where the Free–Wilson part accounts for a set of substituents and the Hansch part for another set.

Another mixed model, called here **Site-Property analysis** (*SP analysis*), can be obtained [Todeschini and Consonni, 2000]; it represents information regarding the presence of each substituent group in each site by the corresponding physico-chemical properties, that is, the information of the indicator variables $I_{i,ks}$ of the Fujita–Ban analysis is preserved in each site but

is represented by the set of selected properties:

$$\hat{y}_i = b_0 + \sum_{s=1}^S \sum_{j=1}^J b_{sj} \cdot \phi_{is,j}$$

where S is the number of substitution sites, J the number of properties, and $\phi_{is,j}$ the j th substituent group property in the s th site for the i th compound, that is,

$$\phi_{is,j} = \sum_{k=1}^{N_s} \phi_{ks,j} \cdot I_{i,ks}$$

where

$$\sum_{k=1}^{N_s} I_{i,ks} = 1$$

Therefore, *SP analysis* can be performed only if all substituent group constants are available for all the substituents in the data set. The total number of variables is $S \times J$. This approach allows complete \rightarrow *reversible decoding*, that is, the possibility to interpret by the model *how* and *where* the response is increased/decreased.


By assuming that a response would depend on both the holistic properties of molecules and the local specific group contributions a mixed **Global-Site-Property analysis** (*GSP analysis*) can be achieved by a generalized model such as

$$\hat{y}_i = b_0 + \sum_{l=1}^{p'} b'_l \cdot \Phi_{il} + \sum_{s=1}^S \sum_{j=1}^J b_{sj} \cdot \phi_{is,j}$$

where Φ are generic global properties, that is, global descriptors obtained by any method, p' the number of selected descriptors, and $\phi_{is,j}$ the j th substituent group property in the s th site for the i th compound. The total number of variables is $S \times J + p'$.

In Hansch analysis and related approaches, the statistical problems due to the relatively high number of variables with respect to the number of compounds have to be faced using \rightarrow *variable selection* procedures.

Although the predictive power of a model is considered to be a criterion for the relevance of QSAR models, the main purpose of Hansch analysis and related approaches such as Free–Wilson analysis concerns not prediction, but a better understanding of the chemical problem.

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

- **Hansch bilinear models** \rightarrow Hansch analysis
- **Hansch descriptors** \rightarrow Hansch analysis
- **Hansch–Free–Wilson mixed models** \rightarrow Hansch analysis
- **Hansch–Fujita hydrophobic substituent constants** \rightarrow lipophilicity descriptors
- **Hansch linear model** \rightarrow Hansch analysis
- **Hansch nonlinear model** \rightarrow Hansch analysis
- **Hansch parabolic model** \rightarrow Hansch analysis
- **Harary–Balaban index** \rightarrow distance matrix

- Harary Cluj detour indices → Harary indices
- Harary Cluj distance indices → Harary indices
- Harary connectivity index → distance matrix
- Harary detour/distance indices → Harary indices
- Harary detour indices → Harary indices
- Harary index → distance matrix
- Harary index → Harary indices

■ Harary indices

The **Harary index** H [Plavšić, Nikolić *et al.*, 1993b], also called **RDSUM index** [Ivanciuc, Balaban *et al.*, 1993b], is a molecular topological index derived from the → reciprocal distance matrix \mathbf{D}^{-1} by the → Wiener operator Wi :

$$H \equiv RDSUM \equiv Wi(\mathbf{D}^{-1}) = \frac{1}{2} \cdot \sum_{i=1}^A \sum_{j=1}^A d_{ij}^{-1} = \frac{1}{2} \sum_{i=1}^A RDS_i \quad j \neq i$$

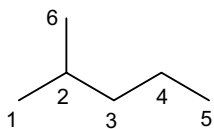
where RDS_i is the → reciprocal distance sum of the i th vertex.

The Harary index increases with both molecular size and → molecular branching; it is therefore a measure of molecular compactness like the → Wiener index. However, the Harary index seems to be a more discriminating index than the Wiener index. A variant H' of the Harary index, called **Harary number**, was derived from the → reciprocal square distance matrix \mathbf{D}^{-2} [Mihalić and Trinajstić, 1992; Plavšić, Nikolić *et al.*, 1993b], still from the Wiener operator Wi :

$$H' \equiv Wi(\mathbf{D}^{-2}) = \frac{1}{2} \cdot \sum_{i=1}^A \sum_{j=1}^A d_{ij}^{-2} \quad j \neq i$$

Example H1

Reciprocal distance matrix \mathbf{D}^{-1} and Harary index H for 2-methylpentane.



$\mathbf{D}^{-1} =$

Atom	1	2	3	4	5	6	RDS_i
1	0	1	0.50	0.33	0.25	0.50	2.58
2	1	0	1	0.50	0.33	1	3.83
3	0.50	1	0	1	0.50	0.50	3.50
4	0.33	0.50	1	0	1	0.33	3.16
5	0.25	0.33	0.50	1	0	0.25	2.33
6	0.50	1	0.50	0.33	0.25	0	2.58

$$H = \frac{1}{2} \cdot \sum_{i=1}^6 \sum_{j=1}^6 [\mathbf{D}^{-1}]_{ij} = \frac{18}{2} = 9$$

By generalization, Harary indices and **hyper-Harary indices** (or **hyper-Harary numbers**) are all → molecular descriptors derived from the application of the Wiener operator to reciprocal → graph-theoretical matrices; Harary indices are obtained from edge-type matrices, whose nonvanishing off-diagonal elements are only those corresponding to pairs of adjacent vertices, while the hyper-Harary indices are calculated from path-type matrices, whose nonvanishing

elements correspond to all pairs of vertices [Diudea, 1997c]. Other topological indices based on a modified reciprocal distance matrix are the \rightarrow *constant interval reciprocal indices*.

The most important Harary indices are listed below.

- **Harary Wiener indices**

The Harary index and hyper-Harary index, defined only for acyclic graphs, are obtained from \rightarrow *reciprocal Wiener matrix* [Diudea, 1997c; Diudea and Gutman, 1998]. The two following indices are derived, respectively, from the reciprocal edge-Wiener matrix \mathbf{W}_e^{-1} and reciprocal path-Wiener matrix \mathbf{W}_p^{-1} :

$$H_{W_e} \equiv Wi(\mathbf{W}_e^{-1}) = \frac{1}{2} \cdot \sum_{i=1}^A \sum_{j=1}^A [\mathbf{W}_e^{-1}]_{ij}$$

$$H_{W_p} \equiv Wi(\mathbf{W}_p^{-1}) = \frac{1}{2} \cdot \sum_{i=1}^A \sum_{j=1}^A [\mathbf{W}_p^{-1}]_{ij}$$

It must be noted that while the indices obtained by applying the Wiener operator to the distance matrix \mathbf{D} and to the \rightarrow *edge-Wiener matrix* \mathbf{W}_e are equal (i.e., the \rightarrow *Wiener index*), the corresponding Harary indices are not, that is, $H \neq H_{W_e}$.

- **hyper-Harary distance index**

This is obtained from the \rightarrow *reciprocal distance-path matrix* \mathbf{D}_p^{-1} as

$$H_{D_p} \equiv Wi(\mathbf{D}_p^{-1}) = \frac{1}{2} \cdot \sum_{i=1}^A \sum_{j=1}^A [\mathbf{D}_p^{-1}]_{ij}$$

For acyclic graphs, the equality between the \rightarrow *hyper-distance-path index* D_p and the \rightarrow *hyper-Wiener index* WW is not true for the corresponding hyper-Harary indices, that is, $H_{D_p} \neq H_{W_p}$.

- **Harary detour indices**

These are obtained from the \rightarrow *reciprocal detour matrix* Δ^{-1} as [Diudea, Katona *et al.*, 1998]

$${}^1H_{\Delta} \equiv Wi(\Delta_e^{-1}) = \frac{1}{2} \cdot \sum_{i=1}^A \sum_{j=1}^A a_{ij} \cdot [\Delta^{-1}]_{ij} \quad \text{and} \quad H_{\Delta} \equiv Wi(\Delta^{-1}) = \frac{1}{2} \cdot \sum_{i=1}^A \sum_{j=1}^A [\Delta^{-1}]_{ij}$$

where Δ_e^{-1} is the reciprocal edge-detour matrix that accounts only for pairs of adjacent vertices and a_{ij} are the elements of the adjacency matrix equal to one for pairs of adjacent vertices and zero otherwise.

- **Harary detour-distance indices**

These are obtained from the \rightarrow *reciprocal detour-distance combined matrix* as

$${}^1H_{\Delta D} = \sum_{i=1}^A \sum_{j=1}^A a_{ij} \cdot [\Delta \wedge \mathbf{D}^{-1}]_{ij} \quad H_{\Delta D} = \sum_{i=1}^A \sum_{j=1}^A [\Delta \wedge \mathbf{D}^{-1}]_{ij}$$

where a_{ij} are the elements of the adjacency matrix equal to one for pairs of adjacent vertices, and zero otherwise. The same index values are, obviously, obtained from the corresponding transpose matrix, that is, the \rightarrow *reciprocal distance-detour combined matrix*.

• Harary Cluj-distance indices

These are molecular indices derived from \rightarrow *reciprocal Cluj matrices*. The Harary-type index is calculated from the reciprocal edge-Cluj-distance matrix \mathbf{CJD}_e^{-1} as

$$H_{CJD_e} \equiv Wi(\mathbf{CJD}_e^{-1}) = \frac{1}{2} \cdot \sum_{i=1}^A \sum_{j=1}^A [\mathbf{CJD}_e^{-1}]_{ij}$$

while the hyper-Harary-type index from the reciprocal path-Cluj-distance matrix \mathbf{CJD}_p^{-1} as

$$H_{CJD_p} \equiv Wi(\mathbf{CJD}_p^{-1}) = \frac{1}{2} \cdot \sum_{i=1}^A \sum_{j=1}^A [\mathbf{CJD}_p^{-1}]_{ij}$$

In acyclic graphs, the Harary edge-Cluj-distance index H_{CJD_e} coincides with the Harary Wiener index H_{W_e} and the Harary Szeged index H_{SZ_e} ($H_{CJD_e} = H_{W_e} = H_{SZ_e}$), for the corresponding hyper-Harary indices the following relationships hold: $H_{CJD_p} = H_{W_p} \neq H_{SZ_p}$; in cyclic graphs, only $H_{CJD_e} = H_{SZ_e}$, while the other indices give distinct values [Diudea, Pârv *et al.*, 1997b].

• Harary Cluj-detour indices

These are other molecular indices derived from \rightarrow *reciprocal Cluj matrices*. The Harary-type index is calculated from the reciprocal edge-Cluj-detour matrix \mathbf{CJA}_e^{-1} as [Diudea, Katona *et al.*, 1998]

$$H_{CJA_e} \equiv Wi(\mathbf{CJA}_e^{-1}) = \frac{1}{2} \cdot \sum_{i=1}^A \sum_{j=1}^A [\mathbf{CJA}_e^{-1}]_{ij}$$

and the hyper-Harary-type index from the reciprocal path-Cluj-detour matrix \mathbf{CJA}_p^{-1} :

$$H_{CJA_p} \equiv Wi(\mathbf{CJA}_p^{-1}) = \frac{1}{2} \cdot \sum_{i=1}^A \sum_{j=1}^A [\mathbf{CJA}_p^{-1}]_{ij}$$

• Harary Szeged indices

These are molecular indices derived from \rightarrow *reciprocal Szeged matrices*. The Harary-type index is obtained from the reciprocal edge-Szeged matrix \mathbf{SZ}_e^{-1} as [Diudea, 1997c]

$$H_{SZ_e} \equiv Wi(\mathbf{SZ}_e^{-1}) = \frac{1}{2} \cdot \sum_{i=1}^A \sum_{j=1}^A [\mathbf{SZ}_e^{-1}]_{ij}$$

and the hyper-Harary-type index from the reciprocal path-Szeged matrix \mathbf{SZ}_p^{-1} :

$$H_{SZ_p} \equiv Wi(\mathbf{SZ}_p^{-1}) = \frac{1}{2} \cdot \sum_{i=1}^A \sum_{j=1}^A [\mathbf{SZ}_p^{-1}]_{ij}$$

In acyclic graphs, the Harary Szeged index H_{SZ_e} coincides with the Harary Wiener index H_{W_e} ($H_{SZ_e} = H_{W_e}$) while the corresponding hyper-Harary indices are different ($H_{SZ_p} \neq H_{W_p}$); in cyclic graphs all these indices differ [Diudea, Pârv *et al.*, 1997b].

• Harary walk indices

The Harary walk indices are obtained from the \rightarrow reciprocal walk matrix $\mathbf{W}_{(\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3)}^{-1}$:

$$H_{W_{(\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3)}} \equiv W_i(\mathbf{W}_{(\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3)}^{-1}) = \frac{1}{2} \cdot \sum_{i=1}^A \sum_{j=1}^A [\mathbf{W}_{(\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3)}^{-1}]_{ij}$$

where \mathbf{M}_1 , \mathbf{M}_2 , and \mathbf{M}_3 are square $A \times A$ matrices [Diudea, 1997c].

📖 [Estrada and Rodriguez, 1997]

- **Harary matrix** \equiv reciprocal distance matrix \rightarrow distance matrix
- **Harary number** \rightarrow distance matrix
- **Harary number** \rightarrow Harary indices
- **Harary Szeged indices** \rightarrow Harary indices
- **Harary walk indices** \rightarrow Harary indices
- **Harary Wiener indices** \rightarrow Harary indices
- **hardness density** \equiv local hardness \rightarrow quantum-chemical descriptors (\odot hardness indices)
- **hardness indices** \rightarrow quantum-chemical descriptors
- **harmonic mean** \rightarrow statistical indices (\odot indices of central tendency)
- **harmonic oscillator model of aromaticity index** \equiv HOMA index \rightarrow delocalization degree indices
- **harmonic oscillator stabilization energy** \equiv HOSE index \rightarrow delocalization degree indices
- **harmonic topological index** \rightarrow vertex degree
- **Hartley information** \rightarrow information content
- **HASA index** \equiv SSAA index \rightarrow charged partial surface area descriptors
- **HASA₂ index** \rightarrow charged partial surface area descriptors (\odot SSAA index)
- **hash structural codes** \rightarrow substructure descriptors

■ Hasse diagram

Among the \rightarrow ranking methods, the Hasse diagram is a graphical means of illustrating partial order ranking proposed by Hasse [Hasse, 1952]. It was introduced in environmental sciences and QSAR/QSPR studies by Halfon [Halfon and Reggiani, 1986; Halfon, 1989] and refined by Brüggemann [Brüggemann and Bartel, 1999; Brüggemann, Bücherl *et al.*, 1999; Brüggemann, Pudenz *et al.*, 2001] and Carlsen [Carlsen, Sørensen *et al.*, 2002]. The first applications on chemical structure descriptors were proposed by Klein [Klein and Babic, 1997] and Ivanciuc [Ivanciuc, Ivanciuc *et al.*, 2000e]. Hasse diagrams were also used to represent \rightarrow DNA sequences.

Given a set Q of n elements, each described by a vector \mathbf{x} of p variables (attributes), the two elements s and t belonging to Q are comparable if for all the variables x_j either $x_j(t) \geq x_j(s)$ or $x_j(s) \geq x_j(t)$. If $x_j(t) \geq x_j(s)$ for all x_j ($j = 1, \dots, p$) then $t \triangleright s$, that is, t covers s (or s is covered by t). The request “for all” is very important and is called the *generality principle*:

$$t \triangleright s \Leftrightarrow x_j(t) \geq x_j(s) \quad \forall j \in \{1, p\}$$

The ordering relationships between all the pairs of elements are collected into the **Hasse matrix**; for each pair of elements s and t the entry of this matrix is

$$[\mathbf{H}]_{st} = \begin{cases} +1 & \text{if } x_j(s) \geq x_j(t) \quad \forall j \in \{1, p\} \\ -1 & \text{if } x_j(t) \geq x_j(s) \quad \forall j \in \{1, p\} \\ 0 & \text{otherwise} \end{cases}$$

In practice, if the entry $s-t$ contains $+1$, the entry $t-s$ contains -1 ; if the entry $s-t$ contains 0 , also the entry $t-s$ contains 0 . Then, the Hasse matrix is a square $n \times n$ antisymmetric matrix, whose elements take only the values 0 and ± 1 . Moreover, in presence of elements having the same variable values (for all the variables), in both the corresponding entries of the Hasse matrix ($s-t$ and $t-s$), a value equal to 1 is stored. In other words, the object t “dominates” the object s if no contradictions are present in all the variables describing the data; otherwise, if for some variables t “dominates” s and for some others s “dominates” t , the two objects are *not comparable*.

Moreover, if only one criterion is used or all the criteria have a \rightarrow Spearman rank correlation equal to one, that is, all the variables provide the same ordering, then a complete or total order is obtained, and all the alternatives are comparable.

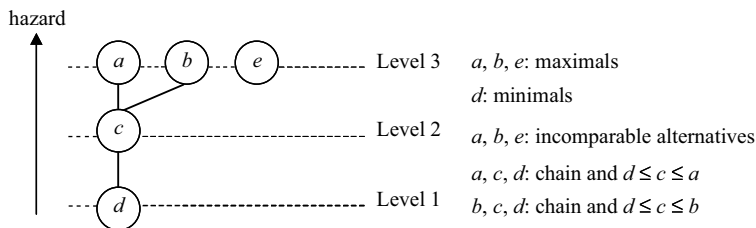
Example H2

Hasse matrix \mathbf{H} and Hasse diagram calculated from five objects described by two variables. Dominance is defined by the maximum values for both variables.

Object	Variable x_1	Variable x_2
a	5	7
b	6	5
c	4	4
d	2	1
e	1	8

$\mathbf{H} =$

	a	b	c	d	e
a	0	0	+1	+1	0
b	0	0	+1	+1	0
c	-1	-1	0	+1	0
d	-1	-1	-1	0	0
e	0	0	0	0	0



- **Hasse matrix** → Hasse diagram
- **HATS indices** → GETAWAY descriptors
- **HATS total index** → GETAWAY descriptors
- **Hausdorff chirality measure** → chirality descriptors
- **HB₁ and HB₂ parameters** → hydrogen-bonding descriptors
- **HBCA index** → charged partial surface area descriptors
- **HB-CPSA descriptors** ≡ *hydrogen-bond charged partial surface area descriptors* → charged partial surface area descriptors
- **H-bonding descriptors** ≡ *hydrogen-bonding descriptors*
- **HB parameter** → hydrogen-bonding descriptors
- **HBSA index** → charged partial surface area descriptors
- **HCD descriptors** → molecular descriptors (⊙ invariance properties of molecular descriptors)
- **HDCA index** → charged partial surface area descriptors
- **HDCA₂ index** → charged partial surface area descriptors
- **H-depleted molecular graph** → molecular graph
- **HDSA index** ≡ *SSAH index* → charged partial surface area descriptors
- **HDSA₂ index** → charged partial surface area descriptors (⊙ HDCA₂ index)
- **Henry's law constant** → physico-chemical properties
- **Hermite-like wave functions** → characteristic polynomial-based descriptors
- **Herndon resonance energy** → delocalization degree indices
- **HE-state fields** → electrotopological state indices
- **HE-state index** ≡ *hydrogen electrotopological state index* → electrotopological state indices
- **heteroatom-corrected extended adjacency matrix** → extended adjacency matrices
- **heteroatom/multiplicity-corrected extended adjacency matrix** → extended adjacency matrices
- **H-filled molecular graph** → molecular graph
- **hierarchical fragment description** → substructure descriptors
- **hierarchically ordered extended connectivities algorithms** → canonical numbering
- **hierarchical QSAR approach** → Structure/Response Correlations
- **higher order map matrices** → biodescriptors (⊙ proteomics maps)
- **higher order Wiener numbers** → Wiener matrix
- **higher order χ matrices** → weighted matrices (⊙ weighted adjacency matrices)
- **highest occupied molecular orbital** → quantum-chemical descriptors
- **highest occupied molecular orbital energy** → quantum-chemical descriptors
- **highest scoring common substructure** → maximum common substructure
- **Higuchi–Davis model** → Hansch analysis

■ Hildebrand solubility parameter (δ_H)

A measure of the intermolecular interactions between solute molecules and their environment, defined as

$$\delta_H = \sqrt{\frac{-E_c}{\bar{V}}} = \sqrt{\frac{\Delta H_v - RT}{\bar{V}}}$$

where E_c is the cohesion energy between liquid molecules defined as a function of → *polarizability*, → *ionization potential*, and → *dipole moment*; \bar{V} is the → *molar volume* of the compound;

ΔH_v the vaporisation enthalpy of the liquid at 298°K, T the absolute temperature; and R the gas universal constant [Hildebrand and Scott, 1950]. For apolar or moderately polar compounds, the vaporization enthalpy can be estimated by their boiling points (bp, °K) as

$$\Delta H_v(\text{cal/mole}) = -2950 + 23.7 \cdot \text{bp} + 0.02 \cdot \text{bp}^2$$

Often referred to as the **solvent cohesive energy density**, the Hildebrand solubility parameter is considered a measure of the solvent contribution to the \rightarrow *cavity term*, and is used as a correction factor in the \rightarrow *solvatochromic equation*. It is related to the general definition of **London cohesive energy** between two interacting species:

$$\epsilon_L = \frac{3 \cdot \alpha_i \cdot \alpha_j}{2 \cdot r_{ij}^6} \cdot \frac{IP_i \cdot IP_j}{IP_i + IP_j}$$

where α and IP are the \rightarrow *polarizability* and the \rightarrow *ionization potential* of the two species, respectively, and r is their \rightarrow *geometric distance*.

Moreover, for large molecules or polymeric systems, a solubility parameter can be calculated from group contributions as [Small, 1953]

$$\delta_s = \frac{\sum_k F_k}{\bar{V}}$$

where F_k is the molar attraction constant of the k th substituent group of the compound and the sum runs over all groups; \bar{V} is the molar volume.

📖 [Kamlet, Carr *et al.*, 1981; Pussemier, De Borger *et al.*, 1989; Mutelet, Ekulu *et al.*, 2002; Stefanis, Constantinou *et al.*, 2004]

- **Hill potential function** \rightarrow molecular interaction fields (⊙ steric interaction fields)
- **H indices** \rightarrow GETAWAY descriptors
- **HINT** \equiv *Hydropatic INteractions* \rightarrow molecular interaction fields (⊙ hydrophobic fields)
- **HLOGP** \rightarrow lipophilicity descriptors
- **HNSO_T index** \rightarrow hydrogen-bonding descriptors
- **H total index** \rightarrow GETAWAY descriptors
- **HOC algorithms** \equiv *hierarchically ordered extended connectivities algorithms* \rightarrow canonical numbering
- **HOC rank descriptors** \rightarrow canonical numbering (⊙ hierarchically ordered extended connectivities algorithms)
- **Hodes statistical-heuristic method** \rightarrow scoring functions
- **Hodgkin similarity index** \rightarrow quantum similarity
- **holograms** \equiv *molecular holograms* \rightarrow substructure descriptors (⊙ fingerprints)
- **Hologram QSAR** \rightarrow substructure descriptors (⊙ fingerprints)
- **holographic vectors** \rightarrow vectorial descriptors
- **Holtz–Stock inductive constant** \rightarrow electronic substituent constants (⊙ inductive electronic constants)
- **HOMA index** \rightarrow delocalization degree indices
- **homeomorphic graphs** \rightarrow graph
- **HOMO–LUMO energy gap** \rightarrow quantum-chemical descriptors

- **HOMO–LUMO energy fraction** → quantum-chemical descriptors
- **HOSE index** → delocalization degree indices
- **Hosoya graph decomposition** → Hosoya Z index
- **Hosoya ID number** → Hosoya Z matrix
- **Hosoya matrix** \equiv *Hosoya Z matrix*
- **Hosoya nonadjacent number** \equiv *nonadjacent number* → Hosoya Z index
- **Hosoya mean information index** → Hosoya Z index
- **Hosoya operator** → characteristic polynomial-based descriptors
- **Hosoya resonance energy** → delocalization degree indices
- **Hosoya total information index** → Hosoya Z index
- **Hosoya-type indices** → characteristic polynomial-based descriptors
- **Hosoya–Wiener index** → double invariants
- **Hosoya–Wiener polynomial** \equiv *Wiener polynomial* → Wiener index
- **Hosoya Z' index** → characteristic polynomial-based descriptors

■ **Hosoya Z index** (\equiv *Z index*)

The Hosoya Z index of a graph G is derived by a combinatorial algorithm and is defined as [Hosoya, 1971]

$$Z = \sum_{k=0}^{\lfloor A/2 \rfloor} a(G, k)$$

where $a(G, k)$, called **nonadjacent number** (or **Hosoya nonadjacent number**), is the number of ways through which k edges may be selected from all of the B edges of graph G , so that no two of them are adjacent, that is, the number of $\rightarrow k$ -*matchings*. A is the number of graph vertices and the Gaussian brackets $\lfloor \]$ represent the greatest integer not exceeding $A/2$. The Z index is calculated by summing the $a(G, k)$ coefficients over all different k values. For any graph, $a(G, 0) = 1$ and $a(G, 1) = B$.

The Hosoya Z index depends on the molecular size as well as on branching and ring closure. It was also found to correlate well with the boiling point.

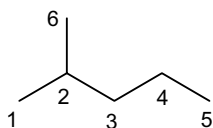
Hosoya graph decomposition is a graph edge decomposition defined as

$$\{\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_j, \dots, \mathcal{E}_{N_k}\}_k$$

where N_k is the nonadjacent number $a(G, k)$ and \mathcal{E}_j is the j th $\rightarrow k$ -*matching* of the graph, that is, a subset of k nonadjacent edges.

Example H3

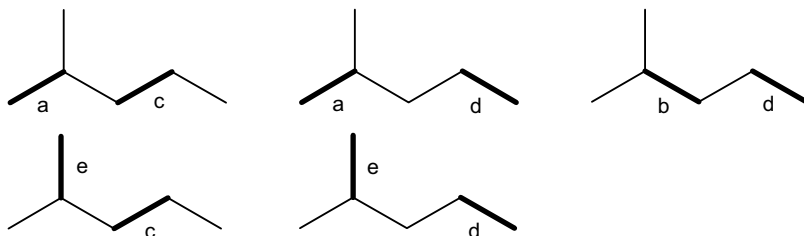
For 2-methylpentane, the H-depleted molecular graph and the nonadjacent numbers are:



$$\begin{aligned} a(G; 0) &= 1 \\ a(G; 1) &= 5 \\ a(G; 2) &= 5 \\ a(G; 3) &= 0 \end{aligned}$$

The number of graph vertices is six, then $[6/2] = 3$, which is the greatest integer not exceeding $6/2 = 3$, and thus $k = 0, 1, 2, 3$.

The computation of $a(G, 2)$ is presented. There are five 2-matchings as shown in the graphs below where the two nonadjacent edges are indicated by bold lines.



The Hosoya graph decomposition for $k = 2$ is

$$E_1 = \{a, c\}, \quad E_2 = \{a, d\}, \quad E_3 = \{b, d\}, \quad E_4 = \{e, c\}, \quad E_5 = \{e, d\}$$

The Hosoya Z index of 2-methylpentane is $Z = 1 + 5 + 5 + 0 = 11$

Hosoya found that the values of the Z index for linear graphs coincide with the \rightarrow *Fibonacci numbers*, that is, $Z = F_A$, where A is the number of vertices in the molecular graph [Hosoya, 1973]; therefore, for a linear graph, the Z index is closely related to the \rightarrow *Merrifield–Simmons index* [Gutman, Hosoya *et al.*, 1992; Randić, Morales *et al.*, 1996].

Using nonadjacent numbers $a(G, k)$ as coefficients, the **Z-counting polynomial** Q of G is a \rightarrow *counting polynomial* defined as

$$Q(G; x) = \sum_{k=0}^{[A/2]} a(G, k) \cdot x^k$$

where the square brackets refer to the largest integer of $A/2$.

The Hosoya Z index can also be defined as the value of the Q polynomial for $x = 1$.

A **modified Hosoya index** (or **Z^* index**) was defined by a generalization of the Z-counting polynomial by treating the powers of x as independent variables [Randić and Zupan, 2001]:

$$Z^*(x_1, x_2, \dots) = \sum_k a(G, k) \cdot x_k$$

where x_k are integer weights representing the number of times each edge has appeared in all disjoint edge patterns.

Closely related to the Z-counting polynomial, the **matching polynomial** (or **acyclic polynomial** or **reference polynomial**) was defined in terms of nonadjacent numbers $a(G, k)$ as [Gutman, Milun *et al.*, 1977; Gutman, 1979; Gutman, Graovac *et al.*, 1982; Hosoya, 1988; Ivanciuc, 1998d; Graovac, Vukicević *et al.*, 2005]

$$M(G; x) = \sum_{k=0}^{[A/2]} (-1)^k a(G, k) \cdot x^{A-2k}$$

For acyclic graphs, the matching polynomial coincides with the \rightarrow graph characteristic polynomial. Moreover, it was demonstrated the following relationship between the Z-counting and matching polynomials [Hosoya, 2003]:

$$M(G; x) = x^A \cdot Q(G; x) \cdot \left(-\frac{1}{x^2}\right)$$

For acyclic graphs the Z-counting polynomial coefficients $a(G, k)$ coincide with the absolute values of the coefficients of the characteristic polynomial of the adjacency matrix (i.e., \rightarrow graph characteristic polynomial) [Nikolić, Plavšić *et al.*, 1992]. Therefore, for any graph, the Hosoya Z index can also be calculated from the matching polynomial coefficients m_{2k} as

$$Z = \sum_{k=0}^{[A/2]} a(G, k) = \sum_{k=0}^{[A/2]} |m_{2k}|$$

and, only for acyclic graphs, from the coefficients c_i of the characteristic polynomial of the adjacency matrix:

$$Z = \sum_{k=0}^{[A/2]} a(G, k) = \sum_{i=0}^A |c_i|$$

By generalization of this last definition of the Hosoya Z index to any graph and any \rightarrow graph-theoretical matrix \mathbf{M} , the \rightarrow Hosoya-type indices were proposed as the sum of the absolute values of the coefficients of the characteristic polynomial of the matrix \mathbf{M} .

Example H4

The Z-counting polynomial of 2-methylpentane is

$$Q(G; x) = 1 + 5x + 5x^2$$

The matching polynomial is

$$M(G; x) = x^6 - 5x^4 + 5x^2$$

The graph characteristic polynomial is

$$Ch(G; x) = x^6 - 5x^4 + 5x^2$$

From the matching polynomial or the characteristic polynomial, the Hosoya Z index is

$$Z = |1| + |-5| + |5| = 11$$

To calculate the Z index for large graphs a composition principle for Z was developed [Hosoya, 1971]: the Z index value of the graph G is obtained as the product of the Z values of graphs G' and G'' , derived from G by cutting an edge, plus the product of the Z values of all graphs derived from G' and G'' by cutting all edges incident to the edge b in the original graph G . The Z value for an empty graph is set at one. It was demonstrated that the Z value for the graph G is uniquely obtained independently of the choice of the edge b to cut in the first step.

The Hosoya Z index for an edge-weighted graph takes into account the edge weights w_b , defining the nonadjacent numbers $a(G, w, k)$ as

$$a(G, w, k) = \sum_{\mathcal{E}_j} \left(\prod_{b=1}^k w_b \right)_{\mathcal{E}_j}$$

where the product is over all edges b of a set \mathcal{E} comprised of k nonadjacent edges and the summation runs over all the subsets of the Hosoya graph decomposition. Obviously, it follows that $a(G, w, 1) = \sum_b w_b$ and, by definition, $a(G, w, 0) = 1$.

The **Hosoya total information index**, denoted as I_Z , is calculated on the Hosoya graph decomposition and is based on the distribution of $a(G, k)$ coefficients. It is an index of \rightarrow *total information content* defined as the following:

$$I_Z = Z \cdot \log_2 Z - \sum_{k=0}^{[A/2]} a(G, k) \cdot \log_2 a(G, k)$$

where Z is the Hosoya index. Analogously, the **Hosoya mean information index** is defined as \rightarrow *mean information content*:

$$\bar{I}_Z = - \sum_{k=0}^{[A/2]} \frac{a(G, k)}{Z} \cdot \log_2 \frac{a(G, k)}{Z}$$

It is noteworthy that I_Z and \bar{I}_Z indices coincide with \rightarrow *information indices on polynomial coefficients* for acyclic graphs.

Example H5

Calculation of Hosoya total and mean information indices is shown for 2-methylpentane. Data from Example H3.

$$I_Z = 11 \times \log_2 11 - 1 \times \log_2 1 - 2 \times (5 \times \log_2 5) = 14.834$$

$$\bar{I}_Z = - \frac{1}{11} \times \log_2 \frac{1}{11} - 2 \times \left(\frac{5}{11} \times \log_2 \frac{5}{11} \right) = 1.349$$

Generalized Hosoya indices Z_m were proposed [Hermann and Zinn, 1995] as counts of nonadjacent molecular paths in the graph G :

$$Z_m = \sum_{k=0}^G a(G, k)_m$$

where the subscript m refers to the order of the index and indicates the path length, $a(G, k)_m$ is the number of all possible combinations of k nonadjacent paths of length m , $a(G, 0) = 1$ by definition, and G is the maximum possible number of k , dependent on the selected path length and the molecule size. The Z_1 index coincides with the original Hosoya Z index and the Z_0 index counts all possible combinations of nonadjacent vertices. These indices must not be confused with the \rightarrow *m Z numbers* based on the sequential erasure of each path from the original graph.

Hosoya Z index was also used to define \rightarrow resonance indices and \rightarrow graphical bond order.

■ Additional references are collected in the thematic bibliography (see Introduction).

■ Hosoya Z matrix

A square symmetric matrix of dimension $A \times A$, A being the number of vertices in the \rightarrow *H-depleted molecular graph* \mathcal{G} . The original Hosoya Z matrix is defined only for acyclic graphs; each off-diagonal element is equal to the \rightarrow *Hosoya Z index* of the subgraph \mathcal{G}' obtained from the graph \mathcal{G} by erasing all edges along the path connecting two vertices v_i and v_j as

$$[Z]_{ij} = \begin{cases} Z(\mathcal{G}') & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

The diagonal entries are zero by definition [Randić, 1994b]. If more than one subgraph is obtained by the erasure procedure, the matrix element is calculated by summing up all of the Hosoya Z indices of the subgraphs.

A general definition of the Hosoya Z matrix (**generalized Hosoya Z matrix**) able to represent both acyclic and cyclic graphs is the following [Plavšić, Šoškić *et al.*, 1997]:

$$[Z]_{ij} = \begin{cases} \frac{\sum_{\min P_{ij}} Z(\mathcal{G}')}{\min P_{ij}} & \text{if } i \neq j \\ Z(\mathcal{G}) & \text{if } i = j \end{cases}$$

where $Z(\mathcal{G}')$ is the Z index of the graph \mathcal{G}' obtained from the graph \mathcal{G} by erasing all edges along the shortest path connecting the vertices v_i and v_j , that is, the \rightarrow *geodesic* $\min P_{ij}$, and the summation goes over all $\min P_{ij}$ geodesics between the considered vertices. The diagonal entries are simply equal to $Z(\mathcal{G})$, that is, the Hosoya Z index of the original graph.

It is interesting to observe that the magnitude of the entries in the matrix \mathbf{Z} decreases as the separation between the vertices increases, it can therefore be expected to simulate the interactions between the pairs of vertices well.

The \mathbf{Z}'/\mathbf{Z} index is among the \rightarrow *graphical bond order descriptors* and can be obtained from the Hosoya Z matrix only by considering the entries relative to pairs of adjacent vertices (i.e., bonds):

$$\frac{Z'}{Z} = \frac{1}{Z} \cdot \text{Wi}(\mathbf{Z}_e) = \frac{1}{2} \cdot \sum_{i=1}^A \sum_{j=1}^A \left(\frac{[Z_e]_{ij}}{Z} \right) = \sum_{b=1}^B \left(\frac{Z(\mathcal{G}')}{Z} \right)_b$$

where Z is the \rightarrow *Hosoya Z index* of the whole graph and Wi is the \rightarrow *Wiener operator* applied to the **edge-Hosoya matrix**, denoted as \mathbf{Z}_e , where the only the nonvanishing elements correspond to pairs of adjacent vertices:

$$\mathbf{Z}_e = \mathbf{Z} \otimes \mathbf{A}$$

\mathbf{A} is the \rightarrow *adjacency matrix*, \otimes the \rightarrow *Hadamard matrix product*, and B is the total number of graph edges. Each term in the sum is a \rightarrow *graphical bond order*.

Other graph invariants derived from the Hosoya Z matrix are the \rightarrow *eigenvalues* and the coefficients of the \rightarrow *characteristic polynomial*. Moreover, sequences of weighted paths and the \rightarrow *weighted path counts* were defined using as the path weights the magnitude of the Z

matrix elements. mZ numbers are calculated as the sum of the magnitude of the entries corresponding to pairs of vertices separated by the shortest path of length m :

$${}^mZ = \sum_{m_{p_{ij}}} [Z]_{ij} = \sum_{m_{p_{ij}}} Z(G - {}^m p_{ij})$$

where the term in the second summation is the Hosoya Z number of the graph G from which the path ${}^m p_{ij}$ is erased. mZ numbers must not be confused with \rightarrow *generalized Hosoya indices* Z_m .

Summing up all mZ numbers, a global molecular descriptor called **Hosoya ID number** ZID is obtained:

$$ZID = A + \sum_m {}^mZ$$

where A is the number of vertices corresponding to 0Z and the summation goes over all path lengths.

From row sums of the Z matrix using the \rightarrow *Ivanciuc–Balaban operator* IB , a Balaban-like index called **L_Z index** is calculated as

$$L_Z \equiv IB(Z) = \frac{B}{C+1} \cdot \sum_{i=1}^{A-1} \sum_{j=i+1}^A a_{ij} \cdot (VS_i(Z) \cdot VS_j(Z))^{-1/2}$$

where B is the number of graph edges, C the \rightarrow *cyclomatic number*, and VS_i and VS_j the row sums of the Hosoya matrix Z corresponding to the vertices v_i and v_j ; a_{ij} are the elements of the \rightarrow *adjacency matrix* which are equal to one for adjacent vertices, and zero otherwise.

Analogously, applying the \rightarrow *Wiener operator* Wi to the matrix Z , a Wiener-type index called **K_Z index** is obtained as

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

📖 [Plavšić, Šoškić *et al.*, 1996a; Janežič, Lučić *et al.*, 2007]

- **Hou fitness function** \rightarrow regression parameters
- **HRNCG index** \rightarrow charged partial surface area descriptors
- **HRNCS index** \rightarrow charged partial surface area descriptors
- **HRPCG index** \rightarrow charged partial surface area descriptors
- **HRPCS index** \rightarrow charged partial surface area descriptors
- **H_1 topological index** \rightarrow connectivity indices
- **Hückel resonance energy** \rightarrow delocalization degree indices
- **Hückel's rule** \rightarrow delocalization degree indices
- **Hurvich–Tsai criterion** \rightarrow regression parameters (☉ Table R1)
- **Hutter likeliness score** \rightarrow scoring functions
- **Hu–Xu ID number** \rightarrow ID numbers
- **Hu–Xu vertex degree** \rightarrow vertex degree
- **hydrated surface area** \rightarrow molecular surface (☉ solvent-accessible molecular surface)

■ Hydration Free Energy Density (HFED)

The empirical hydration free energy density is expressed by a linear combination of some physical properties calculated around the molecule with net atomic charges, polarizabilities, dispersion coefficients of the atoms in the molecule, and solvent accessible surface [Son, Han *et al.*, 1999]. These physical properties are the result of the interaction of the molecule with its environment. To calculate the HFED of a molecule a grid model was proposed; a shell of critical thickness r_C was defined around the solvent-accessible surface with a number of grid points inside (e.g., 8 points/Å³).

The hydration free energy density, denoted by g_k , is calculated at each k th grid point of the shell as

$$g_k = \frac{b_0}{N_G} + \frac{b_1}{N_G} \cdot \sum_{k=1}^{N_G} R_k + b_2 \cdot \left| \sum_{i=1}^A \frac{q_i}{r_{ik}} \right| + b_3 \cdot \sum_{i=1}^A \frac{q_i^2}{r_{ik}^2} + b_4 \cdot \sum_{i=1}^A \frac{\alpha_i}{r_{ik}^3} + b_5 \cdot \sum_{i=1}^A \frac{D_i}{r_{ik}^6}$$

where N_G and A are the numbers of grid points and molecule atoms, respectively; R_k the distance between the center of mass of the molecule and the k th grid point; r_{ik} the distance between the i th atom and the k th grid point; q_i the net atomic charge of the i th atom; and α and D the atomic polarizabilities and the dispersion coefficients. b_0, \dots, b_5 are regression coefficients to be determined by multivariate regression analysis.

The atomic polarizability is the charge dependent effective atomic polarizability (CDEAP) calculated by an empirical method as a linear function of the net atomic charge q_i :

$$\alpha_i = \alpha_i^0 - a_i \cdot q_i$$

where α^0 and a are the effective atomic polarizability of a neutral atom and the charge coefficient, respectively [No, Cho *et al.*, 1993]. The **atomic dispersion coefficient** D is calculated by the Slater–Kirkwood formula [Slater and Kirkwood, 1931]:

$$D_i = \frac{3}{4} \cdot \left(\frac{e \cdot h}{m \cdot \sqrt{e}} \right) \cdot \frac{\alpha_i^2}{\sqrt{\alpha_i / N_{el}}}$$

where h , m , and e are the Planck constant, the mass, and the charge of the electrons; N_{el} is the number of effective electrons of the i th atom.

The hydration free energy ΔG_{HYD} is obtained by summing over the HFED within a threshold distance r_t :

$$\Delta G_{HYD} = \sum_{k=1}^{N_G} g_k$$

where N_G is the number of grid points. The quantity ΔG_{HYD} is the scalar representation of the field around the molecule given by the hydration free energy density; to encode information on the spatial distribution of this physical property the **free energy of hydration density tensor** was also proposed. The elements of the tensor in Cartesian coordinates are defined as

$$\vec{g}_{xx} = \frac{1}{2} \cdot \sum_{k=1}^{N_G} g_k \cdot (2x_k^2 - y_k^2 - z_k^2)$$

$$\vec{g}_{xy} = \frac{3}{2} \cdot \sum_{k=1}^{N_G} g_k \cdot x_k \cdot y_k$$

where \vec{g}_{xx} and \vec{g}_{xy} are xx and xy components of the tensor, and x , y , and z the coordinates of the grid point.

- **hydride group** → molecular graph
- **hydrogen-bond acceptors** → hydrogen-bonding descriptors
- **hydrogen-bond acceptor number** → hydrogen-bonding descriptors
- **hydrogen-bond charged partial surface area descriptors** → charged partial surface area descriptors
- **hydrogen bond acidity** → Linear Solvation Energy Relationships (⊙ hydrogen-bond parameters)
- **hydrogen bond basicity** → Linear Solvation Energy Relationships (⊙ hydrogen-bond parameters)
- **hydrogen-bond donor number** → hydrogen-bonding descriptors
- **hydrogen-bond donors** → hydrogen-bonding descriptors
- **hydrogen-bond electron-acceptor power** \equiv *hydrogen bond basicity* → Linear Solvation Energy Relationships (⊙ hydrogen-bond parameters)
- **hydrogen-bond electron-drawing power** \equiv *hydrogen bond acidity* → Linear Solvation Energy Relationships (⊙ hydrogen-bond parameters)
- **hydrogen-bond index** → hydrogen-bonding descriptors
- **hydrogen-bonding ability constants** → hydrogen-bonding descriptors

■ **hydrogen-bonding descriptors** (\equiv *H-bonding descriptors*)

The hydrogen bond is the bond arising from the interaction between a hydrogen and an electron donor atom, such as oxygen and nitrogen; hydrogen-bonding modifies the electron distribution of the neighbor of the electron-donor atom, thus influencing reactivity. Hydrogen-bonding causes an association of molecules, that is, large aggregates of single molecules. This association influences several → *physico-chemical properties*, such as the compressibility factor, vaporization energy, density, surface tension, parachor, conductivity, dielectric constant, molar refractivity, and boiling and melting points. Moreover, the hydrogen-bonding ability of molecules has long been recognized as being very important in biological reactions, including drug actions.

The theory of hydrogen bonding was fully discussed by Pimentel and McClellan [Pimentel and McClellan, 1960] and Vinogradov and Linnell [Vinogradov and Linnell, 1971].

Intramolecular and intermolecular hydrogen bonds can occur in biological and chemical systems. Moreover, functional groups in the molecule can be distinguished into **Hydrogen-Bond Donor** (HBD) and **Hydrogen-Bond Acceptor** (HBA), the former group having strong electron-withdrawing substituents such as $-\text{OH}$, $-\text{NH}$, $-\text{SH}$, and $-\text{CH}$ and the latter different groups such as $-\text{PO}$, $-\text{SO}$, $-\text{CO}$, $-\text{N}$, $-\text{O}$, $-\text{S}$, and $-\text{F}$; even a π -electron rich system can be considered a H-bond acceptor. Some groups are amphiprotic, that is, with both acceptor and donor ability, such as $-\text{OH}$ and $-\text{NH}$ [Gutmann, 1978].

Hydrogen-bonding ability within a congeneric series of compounds having a common H-bond acceptor or donor can be correlated with the electronic effects of substituents using either

→ *electronic substituent constants* or other physico-chemical properties such as pK_a values. In particular, pK_a is the → *acid dissociation constant*, that is, a measure of the extent of ionization of weakly acid organic compounds. Most approaches for estimating pK_a are based on → *group contribution methods*; other approaches are based on quantum-chemical calculations [Grüber and Buss, 1989; Dixon and Jurs, 1993; Sixt, Altschuh *et al.*, 1996; Schüürmann, Segner *et al.*, 1997; Duboc, 1978; Amić, Davidović-Amić *et al.*, 1999].

Hydrogen-bonding descriptors were introduced in → *Hansch analysis* as well as in → *grid-based QSAR techniques* in the form of → *hydrogen-bonding fields*; moreover, → *hydrogen bond acidity* and → *hydrogen bond basicity* scales, which are among the → *solvatochromic parameters*, were derived both for solutes and solvents by an empirical approach. → *Hydrogen-bond charged partial surface area descriptors* were derived from → *computational chemistry* based on surface areas and partial charges of HBA and HBD atoms or groups.

In analogy with solvatochromic parameters but based on quantum theoretical chemistry, a set of → *quantum-chemical descriptors* intended to describe the hydrogen bonding effects of molecules by → *Theoretical Linear Solvation Energy Relationships* (TLSE) were proposed.

Moreover, H-bond donor ability was estimated by using atomic charge on the most positively charged hydrogen atom in the molecule (Q_H) in conjunction with the → *lowest unoccupied molecular orbital energy* ϵ_{LUMO} ; in an analogous way, H-bond acceptor ability was estimated by using the charge of the most negatively charged atom which is also capable of hydrogen bonding (Q_{MN}) in conjunction with the → *highest occupied molecular orbital energy* ϵ_{HOMO} [Dearden and Ghafourian, 1995; Urrestarazu Ramos, Vaes *et al.*, 1998]. H-bond donor ability was also estimated by using electron → *donor superdelocalizability* S^+ and → *self-atom polarizability* P [Dearden, Cronin *et al.*, 1997].

The other most popular hydrogen-bonding descriptors are listed below. Reviews about hydrogen-bonding parameters are [Hadzi, Kidrić *et al.*, 1990; Dearden and Ghafourian, 1999; Winiwarter, Ax *et al.*, 2003].

• HB parameter

The simplest hydrogen-bonding molecular descriptor, defined as binary variable and accounting for the general ability of the molecule to give hydrogen bonds [Fujita, Nishioka *et al.*, 1977]. A modified HB parameter was proposed [Charton and Charton, 1982] on the basis of the number of hydrogen bonds that a molecule or substituent is capable of forming; for example, HB($-NH_2$) = 2 as a proton donor and HB($-NH_2$) = 1 as a proton acceptor, and HB($-OH$) = 1 as a proton donor and HB($-OH$) = 2 as a proton acceptor.

• I_{HA} and I_{HD} parameters

The simplest hydrogen-bonding substituent descriptors defined as binary variables accounting for the ability of the substituent to give hydrogen bonds [Hansch and Leo, 1979]. I_{HA} is equal to one if the substituent includes at least one H-bond acceptor, otherwise, zero. In the same way, I_{HD} is equal to one if the substituent includes at least one H-bond donor, otherwise, zero.

• Hydrogen-Bond Acceptor number (HBA)

A measure of the hydrogen-bonding ability of a molecule expressed in terms of the number of possible hydrogen-bond acceptors. In particular, it is calculated as the count of lone pairs on oxygen and nitrogen atoms in the molecule.

- **Hydrogen-Bond Donor number (HBD)**

A measure of the hydrogen-bonding ability of a molecule expressed in terms of the number of possible hydrogen-bond donors. In particular, it is calculated as the count of hydrogen atoms bonded to oxygen and nitrogen atoms in the molecule.

📖 [Winiwarter, Bonham *et al.*, 1998]

- **HB₁ and HB₂ parameters**

Substituent descriptors of the hydrogen-bonding ability of functional groups, defined by a set of rules [Yang, Lien *et al.*, 1986].

HB₁ parameter is a count descriptor based on atoms in a group, which possess the ability to form hydrogen bonds. This includes both H-bond acceptors and H-bond donors. The rules for HB₁ are

- oxygen atom is counted as 1, but as zero in $-\text{OCF}_3$;
- nitrogen atom is counted as 1, but as zero if it is bonded to an oxygen atom; moreover, the fragment $\text{N}-\text{N}$ is counted as 1 and the fragment $-\text{N}_3$ as zero;
- hydrogen atom when bonded to oxygen or nitrogen atoms is counted as 1; moreover, it is also counted as 1 in $-\text{C}\equiv\text{C}-\text{H}$ fragment.

For example, HB₁ values for $-\text{NO}$, $-\text{NO}_2$, $-\text{SO}_2\text{NHCH}_3$, and $-\text{CONH}_2$ are 1, 2, 4, and 4, respectively.

The HB₂ parameter is defined as the number of atoms in a group able to form hydrogen bonds multiplied by the value of the strength of hydrogen bond, then divided by 10. The total number of the H-bond acceptors and H-bond donors is calculated by following the rules defined for HB₁. The multiplicative parameters accounting for H-bond strength are 6.05 for oxygen atoms, 5.5 for nitrogen atoms, and 2.5 for hydrogen atoms.

For example, HB₂ values for $-\text{NO}$, $-\text{NO}_2$, $-\text{SO}_2\text{NHCH}_3$, and $-\text{CONH}_2$ are 0.61, 1.21, 2.01, and 1.66, respectively.

📖 [Basak, Niemi *et al.*, 1990b; Basak, 1990]

- **hydrogen-bonding ability constants (I_{H})**

Substituent descriptors defined by measuring the additive contributions of molecular fragments to the hydrogen bonding ability of a molecule [Seiler, 1974]. Such descriptors are calculated from the difference in $\log P$ value in two solvent/water systems:

$$\Delta \log P = \log P_{\text{octanol}} - \log P_{\text{solvent}} = b_0 + \sum_k (I_{\text{H}})_k \cdot N_k$$

where $(I_{\text{H}})_k$ and N_k are the hydrogen-bonding ability constant and the number of occurrences of the k th fragment in the molecule, respectively. b_0 and $(I_{\text{H}})_k$ are regression coefficients estimated by multivariate regression analysis. These substituent descriptors reflect both H-bond donor and H-bond acceptor ability, and are also a function of molecule polarity [Dearden and Ghafourian, 1999].

A set of 23 hydrogen-bonding constants was determined using octanol/water and cyclohexane/water systems; the calculated model was derived from 195 compounds, with intercept $b_0 = -0.16$, $r^2 = 0.935$ and $s = 0.333$. Some I_{H} substituent values are reported in Table H1.

Table H1 Hydrogen-bonding ability constants for some substituent groups.

Substituent	I _H
–N=N–NH– (triazole)	4.24
Aliphatic –COOH	2.88
Aromatic –COOH	2.87
Aromatic –OH	2.60
–CONH–	2.56
–SO ₂ NH–	1.93
Aliphatic –OH	1.82
Aliphatic –NH ₂	1.33
Aromatic –NH ₂	1.18
=N–	1.01
–CO–CH ₂ –CO	0.59
–NR ₁ R ₂ (R ₁ , R ₂ ≠ H)	0.55
–NO ₂	0.45
>C=O	0.31
–C≡N	0.23
–O–	0.11
Ortho-substitution to –OH, –COOH, –NR ₁ R ₂	–0.62

Moreover, $\Delta \log P$ was proposed as a molecular descriptor for modeling hydrogen bonding capacity; it corresponds to the difference between the partition coefficient experimentally determined in octanol/water ($\log P_{ow}$) and partition coefficients determined in other systems, such as octanol/water–heptane/water ($\log P_{alk}$) systems and octanol/water–chloroform/water ($\log P_{CH_3Cl}$) systems [El Tayar, Tsai *et al.*, 1991b; Winiwarter, Ax *et al.*, 2003].

• Raevsky H-bond indices

A set of descriptors characterizing relative H-bond donor and H-bond acceptor abilities of compounds calculated to reproduce the free energy ΔG and enthalpy ΔH of the hydrogen bond complex formation as defined in the thermodynamic equation:

$$\Delta G = \Delta H - T \cdot \Delta S$$

where ΔS is the entropy of complexation and T is the temperature in Kelvin degrees. ΔG was thought of as a multiplicative function of H-bond donor and H-bond acceptor ability as

$$\Delta G = b_0 + b_1 \cdot C_{HD} \cdot C_{HA}$$

where b_0 and b_1 are regression coefficients and C_{HD} and C_{HA} are the free energy H-bond donor and H-bond acceptor factors, respectively [Raevsky, Grigor'ev *et al.*, 1992a; Raevsky, 1997a]. Based on known experimental ΔG values, C_{HD} and C_{HA} values were estimated for 414 and 1298 compounds, respectively, by using the HYBOT program. A value of $C_{HA} = 4.00$ was selected for standard H-bond acceptor (hexamethylphosphoramide) and a value of $C_{HD} = -2.50$ was selected for standard H-bond donor (phenol).

The enthalpy contributions were also estimated and H-bond donor E_{HD} and H-bond acceptor E_{HA} enthalpy factors were also calculated.

Based on these four H-bond factors, Raevsky H-bond indices were therefore proposed as

$$\begin{array}{cccc} C_{\text{HD}}^{\text{max}} & C_{\text{HA}}^{\text{max}} & E_{\text{HD}}^{\text{max}} & E_{\text{HA}}^{\text{max}} \\ \sum C_{\text{HD}} & \sum C_{\text{HA}} & \sum E_{\text{HD}} & \sum E_{\text{HA}} \\ \frac{\sum C_{\text{HD}}}{\text{MW}} & \frac{\sum C_{\text{HA}}}{\text{MW}} & \frac{\sum E_{\text{HD}}}{\text{MW}} & \frac{\sum E_{\text{HA}}}{\text{MW}} \end{array}$$

where the first four descriptors are free energy and enthalpy factors for the strongest H-bond donor atom and H-bond acceptor atom in the molecule; the second set is based on the sums of the free energy and enthalpy factors for all H-bond donor atoms and H-bond acceptor atoms in the molecule; and the third set is constituted by the second set normalized on the molecular weight MW.

📖 [Raevsky, Grigor'ev *et al.*, 1992b, 1993; Schneider, Rüdiger *et al.*, 1993; Raevsky, Dolmatova *et al.*, 1995; Raevsky, 1999; Raevsky, Fetisov *et al.*, 2000; Schaper, Zhang *et al.*, 2001; Raevsky and Skvortsov, 2002]

• Hydrogen-Bond Index (HBI)

An empirical index proposed for chloro–fluoro hydrocarbons (CFC) and defined as [Toropov and Toropova, 2004]

$$HBI = 5000 + N_{\text{H}} - N_{\text{Cl}} - N_{\text{F}}$$

where N_{H} , N_{Cl} , and N_{F} are the number hydrogens, chlorine, and fluorine atoms, respectively; the offset 5000 was added to numerically distinguish this descriptor from other descriptors.

A simple generalization of this index can be proposed taking into account all the halogen atoms in a molecule rather than only fluorine atoms:

$$HBI = 5000 + N_{\text{H}} - N_{\text{Halogens}}$$

• HNSO_T index

A hydrogen-bonding descriptor calculated by adding the total number of lone pairs on oxygen, nitrogen, and sulfur atoms to the number of hydrogen atoms that can be donated by O, N, and S atoms of the molecule in a hydrogen-bonding interaction [MOE – Chemical Computing Group, Inc., 1999; Deretey, Feher *et al.*, 2002].

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

- **hydrogen bonding fields** → molecular interaction fields
- **hydrogen-bond parameters** → Linear Solvation Energy Relationships
- **hydrogen-depleted molecular graph** \equiv *H-depleted molecular graph* → molecular graph
- **hydrogen electrotopological state index** → electrotopological state indices
- **hydrogen-filled molecular graph** \equiv *H-filled molecular graph* → molecular graph
- **hydrogen-included molecular graph** \equiv *H-filled molecular graph* → molecular graph
- **hydropathic atom constants** → molecular interaction fields (⊙ hydrophobic fields)

- **hydropathic interactions** \equiv *Kellogg and Abraham interaction field* \rightarrow molecular interaction fields (\odot hydrophobic fields)
- **hydropathy** \rightarrow molecular interaction fields (\odot hydrophobic fields)
- **hydrophilic effect** \equiv *Moriguchi polar parameter* \rightarrow lipophilicity descriptors

■ hydrophilicity index (H_y)

A simple empirical index related to hydrophilicity of compounds based on \rightarrow *count descriptors* [Todeschini, Vighi *et al.*, 1997]. It is defined as

$$H_y = \frac{(1 + N_{Hy}) \cdot \log_2(1 + N_{Hy}) + N_C \cdot \left(\frac{1}{A} \cdot \log_2 \frac{1}{A}\right) + \sqrt{\frac{N_{Hy}}{A^2}}}{\log_2(1 + A)}$$

where N_{Hy} is the number of hydrophilic groups ($-\text{OH}$, $-\text{SH}$, $-\text{NH}$), N_C the number of carbon atoms, and A the number of atoms (hydrogen excluded). The lowest value of the H_y index is -1 for alkanes with an infinite number of carbon atoms (Table H2).

Table H2 Hydrophilicity values H_y for some compounds.

Compound	N_{Hy}	N_C	A	H_y
2,3,4,5,6-hydroxyphenol	6	6	12	4.881
H_2O_2	2	0	2	3.446
H_2O	1	0	1	3.000
4-OH	4	4	8	3.268
Triols	3	3	6	2.492
Carbonic acid	2	3	6	1.317
Diols	2	2	4	1.769
Methanol	1	1	2	1.262
Ethanol	1	2	3	0.638
Decanediol	2	10	12	0.509
Propanol	1	3	4	0.323
Butanol	1	4	5	0.132
Pentanol	1	5	6	0.004
Methane	0	1	1	0.000
$N_{Hy} = 0$ and $N_C = 0$	0	0	2	0.000
Decanol	1	10	11	-0.294
Ethane	0	2	2	-0.631
Pentane	0	5	5	-0.898
Decane	0	10	10	-0.960
Alkane with $N_C = 1000$	0	1000	1000	-1.000

📖 [Gramatica, Corradi *et al.*, 2000; Yao, Zhang *et al.*, 2002; Put, Perrin *et al.*, 2003; Jelcic, 2004; Stanton, Mattioni *et al.*, 2004; Hancock, Put *et al.*, 2005]

- **hydrophilic-lipophilic balance** \rightarrow grid-based QSAR techniques (\odot VolSurf descriptors)
- **hydrophobic fields** \rightarrow molecular interaction fields

- **hydrophobic fragmental constants** \equiv *Nys-Rekker hydrophobic fragmental constants* \rightarrow lipophilicity descriptors
- **hydrophobicity** \rightarrow lipophilicity descriptors
- **hydrophobic substituent constants** \equiv *Hansch–Fujita hydrophobic substituent constants* \rightarrow lipophilicity descriptors
- **hydropoles** \rightarrow Comparative Molecular Moment Analysis
- **hyper-Cluj-detour index** \rightarrow Cluj matrices
- **hyper-Cluj-distance index** \rightarrow Cluj matrices
- **hyperconjugation effect** \rightarrow electronic substituent constants
- **hyper-detour index** \rightarrow detour matrix
- **hyper-distance-path index** \rightarrow distance-path matrix
- **hyper-Harary distance index** \rightarrow Harary indices
- **hyper-Harary indices** \rightarrow Harary indices
- **hyper-Harary numbers** \equiv *hyper-Harary indices* \rightarrow Harary indices
- **hypermolecule** \rightarrow hyperstructure-based QSAR techniques
- **hyperstructure** \rightarrow hyperstructure-based QSAR techniques

■ hyperstructure-based QSAR techniques

These are QSAR techniques based on the construction of a **hyperstructure** defined as a virtual structure built by overlapping the training set structures such that some atoms and bonds of different structures coincide.

A hyperstructure built by overlapping molecular graphs is called **molecular supergraph** (MSG) and it can be considered as a certain graph such that each training set structure can be represented as its subgraph. A 3D hyperstructure based on the \rightarrow *molecular geometry* of the training set compounds is called **hypermolecule**.

The most important QSAR techniques based on a hyperstructure are \rightarrow *minimal topological difference*, \rightarrow *DARC/PELCO analysis*, and \rightarrow *molecular field topology analysis*.

- **hyper-Szeged index** \rightarrow Szeged matrices
- **hyper-Wiener index** \rightarrow Wiener matrix

■ hyper-Wiener-type indices

These are molecular descriptors calculated from the \rightarrow *H-depleted molecular graph* by analogy with the \rightarrow *hyper-Wiener index*. The general formula, called by Ivanciuc **hyper-Wiener operator** [Ivanciuc, Ivanciuc *et al.*, 1997; Ivanciuc, 2001c], to calculate hyper-Wiener-type indices from vertex- and/or edge-weighted molecular graphs is

$$HyWi(\mathbf{M}; w) = \frac{1}{2} \cdot \sum_{i=1}^A \sum_{j=i}^A ([\mathbf{M}(w)]_{ij}^2 + [\mathbf{M}(w)]_{ij})$$

where A is the number of graph vertices and $\mathbf{M}(w)$ is any square symmetric \rightarrow *graph-theoretical matrix*, calculated with the \rightarrow *weighting scheme* w . Note that, unlike the definition of the original hyper-Wiener index proposed by Klein [Klein, Lukovits *et al.*, 1995], also the diagonal elements of the considered matrix \mathbf{M} , which are usually different from zero in the case of weighted graphs, are taken into account in the hyper-Wiener operator.

If \mathbf{M} is the \rightarrow *distance matrix*, the classical hyper-Wiener index is obtained, $HyWi(\mathbf{D}) = WW$.

An example of the hyper-Wiener-type index is the **Lu index** proposed to describe multiple bond and heteroatom-containing molecules [Lu, Guo *et al.*, 2006a, 2006b, 2006c]. It is calculated from the \rightarrow *bond length-weighted distance matrix* $\mathbf{D}(r^*)$ and defined as

$$Lu \equiv HyWi[\mathbf{D}(r^*)] = A^{1/2} \cdot \log \left[\frac{1}{2} \cdot \sum_{i=1}^A \sum_{j=1}^A [(d_{ij}(r^*))^2 + d_{ij}(r^*)] \right]$$

where $d_{ij}(r^*)$ is the bond length-weighted interatomic distance calculated by adding the relative bond lengths of the edges along the shortest path. The relative bond length r^* is calculated as the ratio of each bond length r_{ij} over the bond length of C–C bond (1.54 Å).

Note that, unlike the definition of the original hyper-Wiener index, a log transformation is applied and the summations are over all the matrix elements and not only on the effective different distances in the graph.

The Lu index was demonstrated to well correlate with boiling point, molar refraction, and gas heat capacity of a number of organic compounds including aliphatic aldehydes and ketones.

Another hyper-Wiener-type index is the \rightarrow *resistance distance hyper-Wiener index*.

➤ **hyper-Wiener operator** \rightarrow hyper-Wiener-type indices