sarchitect

Theory – Descriptors

Molecular descriptors are numerical values obtained by the quantification of various structural and physicochemical characteristics of the molecule. It is envisaged that molecular descriptors quantify these attributes so as to determine the behavior of the molecule and the way the molecule interacts with a physiological system. Since the exact mechanism of drug activity is unknown in many cases, it is desirable to start with descriptors spanning as many attributes of the molecules as possible and then assess their ability to predict the desired activity/property. Sarchitect computes over a thousand descriptors covering constitutional, topological and conformational spaces of compounds.

Terminology

Conventional bond order π_b^* of a bond *b* is 1, 2, 3 and 1.5 for single, double, triple and aromatic bonds respectively.

Molecular Graph. A molecule may be considered as a graph with the atoms as the vertices and bonds as the edges. When a molecular graph is defined without hydrogens, it is called a H-depleted molecular graph.

Terminal Vertices. A vertex in a graph is called terminal if it has only one neighbor in the graph.

Topological distance. This is the shortest distance between a pair of atoms in a graph, defined as the number of edges in the shortest path between the atom pair in the molecular graph. For instance, the topological distance between the two carbons in ethane is 1 and that between the farthest carbons in benzene is 3.

Adjacency Matrix. The adjacency matrix for a molecule with *n* atoms is a $n \times n$ matrix in which the ij^{th} entry is 1 if the atoms *i* and *j* are connected by a bond and 0 otherwise.

Geometric Matrix *G* is defined as a square matrix whose ij^{th} entry is the Euclidean distance between the ith and j^{th} atoms.

Multigraph distance matrix *D** is a weighted distance matrix given by

$$D_{ij}^* = \sum_b 1/\pi_b^*$$

where the sum runs through the sequence of bonds (edges) along the shortest path between atoms *i* and *j*.

Multigraph distance degree σ_i^* of the ith atom is defined as sum of the corresponding row in the multigraph distance matrix :

$$\sigma_i^* = \sum_{j=1}^{n} D_{ij}^*$$

Topological Distance/Level Matrix. The Distance/Level Matrix is a $n \times n$ matrix in which the ij^{th} entry is the topological distance between the atoms *i* and *j*.

Vertex degree δ_i of an atom *i* is defined as the number of atoms adjacent to it in the H-depleted molecular graph.

Constitutional descriptors

Counts

Atom Counts

numAtoms	Number of atoms
nSK	Number of non-H atoms
NH	Number of Hydrogen atoms
NC	Number of Carbon atoms
NN	Number of Nitrogen atoms
NO	Number of Oxygen atoms
NP	Number of Phosphorous atoms
NS	Number of Sulfur atoms
NF	number of Fluorine atoms
nCl	number of chlorine atoms
nBr	number of bromine atoms
nI	number of Iodine atoms
NB	number of Boron atoms
nHM	number of heavy atoms
NX	number of halogen atoms

Bond Counts

numBonds	Number of bonds
nBO	Number of non-H bonds
nCIC	Number of rings
nCIR	Number of circuits
nBM	Number of multiple bonds
SCBO	sum of conventional bond orders
RBN	Number of rotatable bonds
RBF	rotatable bond fraction
nSB	Number of single bonds

nDB	Number of double bonds
nTB	Number of triple bonds
nAB	Number of aromatic bonds

Ring Counts

NR03	Number of 3-membered rings
NR04	Number of 4-membered rings
NR05	Number of 5-membered rings
NR06	Number of 6-membered rings
NR07	Number of 7-membered rings
NR08	Number of 8-membered rings
NR09	Number of 9-membered rings
NR10	Number of 10-membered rings
NR11	number of 11-membered rings
NR12	number of 12-membered rings
NBNZ	number of benzene like rings

Atom centered fragments

(1) The following atom centered fragment descriptors (Viswanadhan, 1989; Viswanadhan, 1993) are defined by looking at the first neighbors of carbon atoms. The neighbors of a carbon atom can be Hydrogens (represented as H), Carbons (represented as R) and heteroatoms (represented as X) in various combinations. = and # indicate double and triple bonds respectively.

CH3R
CH2R2
CHR3
CH3X
CH2RX
CH2X2
CHR2X
CHRX2
CHX3
CR4
CR3X
CR2X2
CRX3
CX4
=CH2
=CHR
=CR2
=CHX

=CRX
=CX2
#CH
#CR
#CX

(2) The following atom centered fragment descriptors are defined for each ring atom that has three neighbors. The typical formats are *A* - - *BC* - - *D*, *A* – *B* (=*C*) – *D* and *A* – *BC* where the atom *C* on a ring is viewed as the center with *A* and *D* being its ring neighbors and *B* (that is not on the same ring as *A*, *C* and *D*) is connected to *B*. The atoms *A*, *C* and *D* can be Hydrogens (represented as H), Carbons (represented as R) and heteroatoms (represented as X). '-', -', '=' and '#' stand for a single, aromatic, double and triple bonds respectively. For example, *R* - - *CH* - - *R* can be defined as a central Carbon atom (C) on an aromatic ring that has two carbon neighbors (R) on the same aromatic ring and the third neighbor outside this ring is a Hydrogen (H).

R-CH-R
R-CR-R
R-CX-R
R-CH-X
R-CR-X
R-CX-X
X-CH-X
X-CR-X
X-CX-X
R-C(=X)-X
X-C(=X)-X

Functional group Counts

nCp number of total primary Carbons (sp3)
nCs number of total secondary Carbons (sp3)
nCt number of total tertiary Carbons (sp3)
nCq number of total quaternary Carbons (sp3)

Property based descriptors

Physical

[MW] molecular weight
[AMW] average molecular weight
[Sv] sum of atomic van der Waals volumes (scaled on Carbon atom)
[Mv] mean atomic van der Waals volume (scaled on Carbon atom)
[Hy] hydrophilic factor 1 (empirical descriptors)

Hydrophilicity index (HyOld)

It is an empirical index related to Hydrophilicity of compounds (Todeschini, 2000).
It is defined as:

$$H_y = ((1+N_{Hy})\log_2(1+N_{Hy}) + N_C(1/A.\log_2 1/A + \sqrt{N_{Hy}/A^2})/\log_2(1+A),$$

where, N_{Hy} is the number of hydrophilic groups (-OH, -SH, -NH), N_C is the number of carbon atoms, and A the number of atoms (hydrogen excluded)

HyOld has values between -1 and 3.64.

Electronic

[Se] sum of atomic Sanderson electronegativities (scaled on Carbon atom)
[Sp] sum of atomic polarizabilities (scaled on Carbon atom)
[Me] mean atomic Sanderson electronegativity (scaled on Carbon atom)
[Mp] mean atomic polarizability (scaled on Carbon atom)

mlogP

mLOGP is the Moriguchi octanol-water partition coefficient (logP) calculated using the model developed by Moriguchi et al. (Moriguchi, 1992; Moriguchi, 1994).

mlogP is defined as:

$$\log P = -1.014 + 1.244.(F_{CX})^{0.6} - 1.017.(N_O + N_N)^{0.9} + 0.406.F_{PRX} - 0.145.N_{UNS}^{0.8} + 0.511.I_{HB} + 0.268.N_{POL} - 2.215.F_{AMP} + 0.912.I_{ALK} - 0.392.I_{RNG} 3.684.F_{QN} + 0.474.N_{NO2} + 1.582.F_{NCS} + 0.773.I_{\beta L}$$

Further, the 13 structural parameters comprising the above mentioned regression equation have also been listed as independent descriptors by themselves, namely:

[MLOGP] Moriguchi octanol-water partition coefficient (logP)
[mlogpCX] Moriguchi based lipophilicity descriptor (carbon and halogen atoms)
[mlogpON] Moriguchi based lipophilicity descriptor (nitrogen and oxygen atoms)
[mlogpPRX] Moriguchi based lipophilicity descriptor (proximity efffect)
[mlogpUB] Moriguchi based lipophilicity descriptor (unsaturated bonds)
[mlogpHB] Moriguchi based lipophilicity descriptor (intramolecular H bonds)
[mlogpPOL] Moriguchi based lipophilicity descriptor (Polar substituents)
[mlogpAMP] Moriguchi based lipophilicity descriptor (amphotericity)
[mlogpALK] Moriguchi based lipophilicity descriptor (alkanes and alkenes)
[mlogpRNG] Moriguchi based lipophilicity descriptor (ring structures)
[mlogpQN] Moriguchi based lipophilicity descriptor (quarternary nitrogen)
[mlogpNO2] Moriguchi based lipophilicity descriptor (Nitro groups)

[mlogpNCS] Moriguchi based lipophilicity descriptor (NCS)
[mlogpBLM] Moriguchi based lipophilicity descriptor (beta lactam)

Pharmacophoric Features

[nHDon] number of hydrogen-bond donors
[nHAcc] number of hydrogen-bond acceptors
[a_hyd] number of hydrophobic atoms

Topological (2D) descriptors

BCUT indices

The BCUT (Burden - CAS - University of Texas eigen values) descriptors are the eigen values of a modified connectivity matrix known as the Burden matrix [Burden, 1989].

The Burden matrices *M* are defined such that :

The diagonal elements M_{ww} are the weights w_i for atom A_i where the weights may be some property associated with the atoms such as *m* (relative atomic mass), *p* (polarizability), *e* (Sanderson electronegativity) and *v* (Van der Waals volume).

The non-diagonal elements M_{wk} are 1 if $k = d_{ij}$ and 0 otherwise, where *k* is the lag defined as the topological distance *d* between the atom pair *i-j* and may have a value between 0-8. Thus for a given *k*, the non-diagonal element M_{ij} will be unity if the atoms *i* and *j* are apart by a topological distance *k* and zero otherwise.

For a given *w* and *k*, there are two BCUT descriptors BEHwk and BELwk – BEHwk is the highest positive eigen value of the matrix M_{wk} and will be zero if there are no positive eigen values. BELwk is the lowest negative eigen value of the matrix M_{wk} and zero if there are no negative eigen values [Burden 1997; Pearlman & Smith 1998, 1999; Benigni 1999].

There are 64 BCUT descriptors in sarchitect:

Mass weighted

BEHm1, BEHm2, BEHm3, BEHm4, BEHm5, BEHm6, BEHm7, BEHm8

BELm1, BELm2, BELm3, BELm4, BELm5, BELm6, BELm7, BELm8

van der Waals volume

BEHv1, BEHv2, BEHv3, BEHv4, BEHv5, BEHv6, BEHv7, BEHv8

BELv1, BELv2, BELv3, BELv4, BELv5, BELv6, BELv7, BELv8

Electronegativity

BEHe1, BEHe2, BEHe3, BEHe4, BEHe5, BEHe6, BEHe7, BEHe8

BELe1, BELe2, BELe3, BELe4, BELe5, BELe6, BELe7, BELe8

Polarizability

BEHp1, BEHp2, BEHp3, BEHp4, BEHp5, BEHp6, BEHp7, BEHp8

BELp1, BELp2, BELp3, BELp4, BELp5, BELp6, BELp7, BELp8

Relative mass is defined as the ratio of atomic mass of an atom to that of carbon. Similarly, the other three weights *p*, *e* and *v* are scaled by the corresponding values for Carbon.

where $\varpi \in \{m, v, e, p\}$ and the lag $k \in \{1 \ldots 8\}$ as in Autocorrelation descriptors.

Autocorrelation descriptors.

2D Auto Correlation Indices

The following Autocorrelation descriptors are available in sarchitect. Broto-Moreau Autocorrelation Descriptors (labeled as ATS) Moran Autocorrelation Descriptors (labeled as MATS) Geary Autocorrelation Descriptors (labeled as GATS)

The symbol for each of the autocorrelation descriptors is followed by two indices *d* and *w* where *d* stands for the lag and *w* stands for the weight. Thus, for example, ATS4m means the Broto-Moreau Autocorrelation Descriptor of *lag* 4 that is weighted by *mass*.

The lag is defined as the topological distance *d* between pairs of atoms. The topological distance between a pair of atoms (*i, j*) is given in the ij^{th} entry in the Topological Level Matrix. The lag can have any value from the set {0, 1, 2, 3, 4, 5, 6, 7, 8}.

The weight can be *m* (relative atomic mass), *p* (polarizability), *e* (Sanderson electronegativity) and *v* (Van der Waals volume). *Relative mass* is defined as the ratio of atomic mass of an atom to that of carbon. Similarly, the other three weights *p*, *e* and *v* are scaled by the corresponding values for Carbon.

Let *n* be the number of atoms in the molecule. For any chosen value for lag *d* and any chosen weight *w*, we compute the Autocorrelation Descriptors using the following formulae.

Broto-Moreau Autocorrelation Descriptors

$$ATSdw = \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_{ij} (w_i w_j),$$

where, w_i and w_j are the weights of the atoms *i* and *j*, $\varpi \in \{m, p, e, v\}$, and δ_{ij} is Kronecker delta, that is, $\delta_{ij} = 1$ if the ij^{th} entry in the Topological Level Matrix is = *d,* and $\delta_{ij} = 0$ otherwise.

(Moreau & Broto, 1980a; Moreau & Broto, 1980b; Broto, 1984a; Broto, 1984b; Broto & Devillers, 1990)

[ATS1m] Broto-Moreau autocorrelation of a topological structure - lag 1 / weighted by atomic masses
[ATS2m] Broto-Moreau autocorrelation of a topological structure - lag 2 / weighted by atomic masses

[ATS3m] Broto-Moreau autocorrelation of a topological structure - lag 3 / weighted by atomic masses
[ATS4m] Broto-Moreau autocorrelation of a topological structure - lag 4 / weighted by atomic masses
[ATS5m] Broto-Moreau autocorrelation of a topological structure - lag 5 / weighted by atomic masses
[ATS6m] Broto-Moreau autocorrelation of a topological structure - lag 6 / weighted by atomic masses
[ATS7m] Broto-Moreau autocorrelation of a topological structure - lag 7 / weighted by atomic masses
[ATS8m] Broto-Moreau autocorrelation of a topological structure - lag 8 / weighted by atomic masses
[ATS1v] Broto-Moreau autocorrelation of a topological structure - lag 1 / weighted by atomic van der Waals volumes
[ATS2v] Broto-Moreau autocorrelation of a topological structure - lag 2 / weighted by atomic van der Waals volumes
[ATS3v] Broto-Moreau autocorrelation of a topological structure - lag 3 / weighted by atomic van der Waals volumes
[ATS4v] Broto-Moreau autocorrelation of a topological structure - lag 4 / weighted by atomic van der Waals volumes
[ATS5v] Broto-Moreau autocorrelation of a topological structure - lag 5 / weighted by atomic van der Waals volumes
[ATS6v] Broto-Moreau autocorrelation of a topological structure - lag 6 / weighted by atomic van der Waals volumes
[ATS7v] Broto-Moreau autocorrelation of a topological structure - lag 7 / weighted by atomic van der Waals volumes
[ATS8v] Broto-Moreau autocorrelation of a topological structure - lag 8 / weighted by atomic van der Waals volumes
[ATS1e] Broto-Moreau autocorrelation of a topological structure - lag 1 / weighted by atomic Sanderson electronegativities
[ATS2e] Broto-Moreau autocorrelation of a topological structure - lag 2 / weighted by atomic Sanderson electronegativities
[ATS3e] Broto-Moreau autocorrelation of a topological structure - lag 3 / weighted by atomic Sanderson electronegativities
[ATS4e] Broto-Moreau autocorrelation of a topological structure - lag 4 / weighted by atomic Sanderson electronegativities
[ATS5e] Broto-Moreau autocorrelation of a topological structure - lag 5 / weighted by atomic Sanderson electronegativities
[ATS6e] Broto-Moreau autocorrelation of a topological structure - lag 6 / weighted by atomic Sanderson electronegativities
[ATS7e] Broto-Moreau autocorrelation of a topological structure - lag 7 / weighted by atomic Sanderson electronegativities
[ATS8e] Broto-Moreau autocorrelation of a topological structure - lag 8 / weighted by atomic Sanderson electronegativities
[ATS1p] Broto-Moreau autocorrelation of a topological structure - lag 1 / weighted by atomic polarizabilities
[ATS2p] Broto-Moreau autocorrelation of a topological structure - lag 2 / weighted by atomic polarizabilities
[ATS3p] Broto-Moreau autocorrelation of a topological structure - lag 3 / weighted by atomic polarizabilities
[ATS4p] Broto-Moreau autocorrelation of a topological structure - lag 4 / weighted by atomic polarizabilities
[ATS5p] Broto-Moreau autocorrelation of a topological structure - lag 5 / weighted by atomic polarizabilities
[ATS6p] Broto-Moreau autocorrelation of a topological structure - lag 6 / weighted by atomic polarizabilities
[ATS7p] Broto-Moreau autocorrelation of a topological structure - lag 7 / weighted by atomic polarizabilities
[ATS8p] Broto-Moreau autocorrelation of a topological structure - lag 8 / weighted by atomic polarizabilities

Moran Autocorrelation Descriptors:

$$MATSdw = (n)(A)/(B),$$

where,

$$A = (1/\Delta)(\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_{ij}(\varpi_i - \varpi)(\varpi_j - \varpi)),$$

and

$$B = \sum_{i=1}^{n}(\varpi_i - \varpi)^2,$$

where,

w_i and w_j are the weights of the atoms i and j, $\varpi \in \{m, p, e, v\}$, $\overline{\varpi}$ is the mean of w_i over the entire molecule, and δ_{ij} is Kronecker delta, that is, $\delta_{ij} = 1$ if the ij^{th} entry in the Topological Level Matrix is $= d$, and $\delta_{ij} = 0$ otherwise (Moran, 1950).

[MATS1m] Moran autocorrelation - lag 1 / weighted by atomic masses
[MATS2m] Moran autocorrelation - lag 2 / weighted by atomic masses
[MATS3m] Moran autocorrelation - lag 3 / weighted by atomic masses
[MATS4m] Moran autocorrelation - lag 4 / weighted by atomic masses
[MATS5m] Moran autocorrelation - lag 5 / weighted by atomic masses
[MATS6m] Moran autocorrelation - lag 6 / weighted by atomic masses
[MATS7m] Moran autocorrelation - lag 7 / weighted by atomic masses
[MATS8m] Moran autocorrelation - lag 8 / weighted by atomic masses
[MATS1v] Moran autocorrelation - lag 1 / weighted by atomic van der Waals volumes
[MATS2v] Moran autocorrelation - lag 2 / weighted by atomic van der Waals volumes
[MATS3v] Moran autocorrelation - lag 3 / weighted by atomic van der Waals volumes
[MATS4v] Moran autocorrelation - lag 4 / weighted by atomic van der Waals volumes
[MATS5v] Moran autocorrelation - lag 5 / weighted by atomic van der Waals volumes
[MATS6v] Moran autocorrelation - lag 6 / weighted by atomic van der Waals volumes
[MATS7v] Moran autocorrelation - lag 7 / weighted by atomic van der Waals volumes
[MATS8v] Moran autocorrelation - lag 8 / weighted by atomic van der Waals volumes
[MATS1e] Moran autocorrelation - lag 1 / weighted by atomic Sanderson electronegativities
[MATS2e] Moran autocorrelation - lag 2 / weighted by atomic Sanderson electronegativities
[MATS3e] Moran autocorrelation - lag 3 / weighted by atomic Sanderson electronegativities
[MATS4e] Moran autocorrelation - lag 4 / weighted by atomic Sanderson electronegativities
[MATS5e] Moran autocorrelation - lag 5 / weighted by atomic Sanderson electronegativities
[MATS6e] Moran autocorrelation - lag 6 / weighted by atomic Sanderson electronegativities
[MATS7e] Moran autocorrelation - lag 7 / weighted by atomic Sanderson electronegativities
[MATS8e] Moran autocorrelation - lag 8 / weighted by atomic Sanderson electronegativities
[MATS1p] Moran autocorrelation - lag 1 / weighted by atomic polarizabilities
[MATS2p] Moran autocorrelation - lag 2 / weighted by atomic polarizabilities
[MATS3p] Moran autocorrelation - lag 3 / weighted by atomic polarizabilities
[MATS4p] Moran autocorrelation - lag 4 / weighted by atomic polarizabilities
[MATS5p] Moran autocorrelation - lag 5 / weighted by atomic polarizabilities
[MATS6p] Moran autocorrelation - lag 6 / weighted by atomic polarizabilities
[MATS7p] Moran autocorrelation - lag 7 / weighted by atomic polarizabilities
[MATS8p] Moran autocorrelation - lag 8 / weighted by atomic polarizabilities

Geary Autocorrelation Descriptors

$$GATSdw = ((n-1)/2)(A)/(B),$$

where,

$$A = (1/\Delta)(\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_{ij}(w_i - w_j)^2),$$

and

$$B = \sum_{i=1}^{n}(w_i - \varpi)^2,$$

where, w_i and w_j are the weights of the atoms i and j, $\varpi \in \{m, v, e, p\}$, ϖ is the mean of w_i over the entire molecule, and δ_{ij} is Kronecker delta, that is, δ_{ij} =1 if the ij^{th} entry in the Topological Level Matrix is = d, and δ_{ij} =0 otherwise (Geary, 1954).

Given below is the list of Autocorrelation descriptors implemented in sarchitect:

[GATS1m] Geary autocorrelation - lag 1 / weighted by atomic masses
[GATS2m] Geary autocorrelation - lag 2 / weighted by atomic masses
[GATS3m] Geary autocorrelation - lag 3 / weighted by atomic masses
[GATS4m] Geary autocorrelation - lag 4 / weighted by atomic masses
[GATS5m] Geary autocorrelation - lag 5 / weighted by atomic masses
[GATS6m] Geary autocorrelation - lag 6 / weighted by atomic masses
[GATS7m] Geary autocorrelation - lag 7 / weighted by atomic masses
[GATS8m] Geary autocorrelation - lag 8 / weighted by atomic masses
[GATS1v] Geary autocorrelation - lag 1 / weighted by atomic van der Waals volumes
[GATS2v] Geary autocorrelation - lag 2 / weighted by atomic van der Waals volumes
[GATS3v] Geary autocorrelation - lag 3 / weighted by atomic van der Waals volumes
[GATS4v] Geary autocorrelation - lag 4 / weighted by atomic van der Waals volumes
[GATS5v] Geary autocorrelation - lag 5 / weighted by atomic van der Waals volumes
[GATS6v] Geary autocorrelation - lag 6 / weighted by atomic van der Waals volumes
[GATS7v] Geary autocorrelation - lag 7 / weighted by atomic van der Waals volumes
[GATS8v] Geary autocorrelation - lag 8 / weighted by atomic van der Waals volumes
[GATS1e] Geary autocorrelation - lag 1 / weighted by atomic Sanderson electronegativities
[GATS2e] Geary autocorrelation - lag 2 / weighted by atomic Sanderson electronegativities
[GATS3e] Geary autocorrelation - lag 3 / weighted by atomic Sanderson electronegativities
[GATS4e] Geary autocorrelation - lag 4 / weighted by atomic Sanderson electronegativities
[GATS5e] Geary autocorrelation - lag 5 / weighted by atomic Sanderson electronegativities
[GATS6e] Geary autocorrelation - lag 6 / weighted by atomic Sanderson electronegativities
[GATS7e] Geary autocorrelation - lag 7 / weighted by atomic Sanderson electronegativities
[GATS8e] Geary autocorrelation - lag 8 / weighted by atomic Sanderson electronegativities
[GATS1p] Geary autocorrelation - lag 1 / weighted by atomic polarizabilities
[GATS2p] Geary autocorrelation - lag 2 / weighted by atomic polarizabilities
[GATS3p] Geary autocorrelation - lag 3 / weighted by atomic polarizabilities
[GATS4p] Geary autocorrelation - lag 4 / weighted by atomic polarizabilities
[GATS5p] Geary autocorrelation - lag 5 / weighted by atomic polarizabilities

[GATS6p] Geary autocorrelation - lag 6 / weighted by atomic polarizabilities
[GATS7p] Geary autocorrelation - lag 7 / weighted by atomic polarizabilities
[GATS8p] Geary autocorrelation - lag 8 / weighted by atomic polarizabilities

Topological Indices

Topological Distances

Let $\{x_1 \ldots x_k\}$ be the atoms of a particular chemical element x in the molecule. Similarly, $\{y_1 \ldots y_m\}$ are the atoms of a particular chemical element y in the molecule. We permit the case of $x = y$ as well. Then the Atom Pair Topological Distance between the chemical elements x and y is given by:

$$1/2\left(\sum_{i=1}^{k}\sum_{j=1}^{m} d_{ij}\right)$$

where,

d_{ij} is the topological distance (see Appendix A) between x_i and y_i (Bonchev, 1991; Trinajstic, 1992). Following are the pairs of chemical elements for which the Atom Pair Topological Distances are computed:

[T(N..N)] sum of topological distances between N..N
[T(N..O)] sum of topological distances between N..O
[T(N..S)] sum of topological distances between N..S
[T(N..P)] sum of topological distances between N..P
[T(N..F)] sum of topological distances between N..F
[T(N..Cl)] sum of topological distances between N..Cl
[T(N..Br)] sum of topological distances between N..Br
[T(N..I)] sum of topological distances between N..I
[T(O..O)] sum of topological distances between O..O
[T(O..S)] sum of topological distances between O..S
[T(O..P)] sum of topological distances between O..P
[T(O..F)] sum of topological distances between O..F
[T(O..Cl)] sum of topological distances between O..Cl
[T(O..Br)] sum of topological distances between O..Br
[T(O..I)] sum of topological distances between O..I
[T(S..S)] sum of topological distances between S..S
[T(S..P)] sum of topological distances between S..P
[T(S..F)] sum of topological distances between S..F
[T(S..Cl)] sum of topological distances between S..Cl
[T(S..Br)] sum of topological distances between S..Br
[T(S..I)] sum of topological distances between S..I
[T(P..P)] sum of topological distances between P..P

[T(P..F)] sum of topological distances between P..F
[T(P..Cl)] sum of topological distances between P..Cl
[T(P..Br)] sum of topological distances between P..Br
[T(P..I)] sum of topological distances between P..I
[T(F..F)] sum of topological distances between F..F
[T(F..Cl)] sum of topological distances between F..Cl
[T(F..Br)] sum of topological distances between F..Br
[T(F..I)] sum of topological distances between F..I
[T(Cl..Cl)] sum of topological distances between Cl..Cl
[T(Cl..Br)] sum of topological distances between Cl..Br
[T(Cl..I)] sum of topological distances between Cl..I
[T(Br..Br)] sum of topological distances between Br..Br
[T(Br..I)] sum of topological distances between Br..I
[T(I..I)] sum of topological distances between I..I

Connectivity indices

Kier-Hall Connectivity Indices

The following Kier-Hall Connectivity Indices are calculated from the Hydrogen-depleted molecular graph (Kier & Hall, 1986).

1. Connectivity index Chi-0 through Chi-5
2. Average Connectivity index Chi-0 through Chi-5
3. Valence Connectivity index Chi-0 through Chi-5
4. Average Valence Connectivity index Chi-0 through Chi-5

Connectivity indices Chi-0 through Chi-5 are defined as follows.

χ^0 : Connectivity index Chi-0 is defined as:

$$\chi^0 = \sum_{i=1}^{n} \delta_i^{-1/2},$$

where,

n is the number of nodes in the Hydrogen-depleted graph,

δ_i is the vertex degree of the ith atom defined as the number of non-Hydrogen neighbours in the molecular graph.

The Average Connectivity index Chi-0 is:

$$\chi^0 A = \chi^0 / n$$

χ^1 : Connectivity index Chi-1 is defined as:

$$\chi^1 = \sum_b (\delta_i \delta_j)^{-1/2},$$

where,

b is the number of bonds, the sum runs through all bond in the Hydrogen-depleted molecule, and for each bond $\delta_i \delta_j$ is the product of the vertex degrees of the end atoms *i* and *j*.

The Average Connectivity index Chi-1 is

$$\chi^1 A = \chi^1 / b.$$

Higher Indices: Connectivity indices Chi-m for $2 \le m \le 5$ is defined as:

$$\chi^m = \sum_{k=1}^{K} (\prod \delta_i)_k^{-1/2},$$

where,

(II $\delta_i)_K$ is the product of the vertex degrees of the atoms that form a connected subgraph with *m* edges, and

K is the total number of such distinct connected sub graphs (the H-depleted molecular graph) each having *m* edges.

For any *m,* $0 \le m \le 5$, if we replace the vertex degree δ_i by the **valence vertex degree** δ_i^v

for each atom *i* in the Connectivity index Chi-m, then we get Valence Connectivity Indices Chi-m (Kier & Hall, 1981; Kier & Hall, 1983). That is,

$$\chi^m = \sum_{k=1}^{K} (\prod \delta_i^v)_k^{-1/2},$$

where,

$(\prod \delta_i^v)_k$ is the product of the valence vertex degrees of the atoms that form a connected subgraph with *m* edges, and

K is the total number of such distinct connected subgraphs (the H-depleted molecular graph) each having *m* edges.

The Valence Connectivity Indices account for the presence of heteroatoms and double and triple bonds.

The Average Valence Connectivity index Chi-1 is defined similarly:

$$\chi^{m}v = \sum_{k=1}^{K} (\prod \delta_i^v)_k^{-1/2},$$

[X0] connectivity index chi-0
[X1] connectivity index chi-1 (Randic connectivity index)
[X2] connectivity index chi-2
[X3] connectivity index chi-3
[X4] connectivity index chi-4
[X5] connectivity index chi-5
[X0A] average connectivity index chi-0
[X1A] average connectivity index chi-1
[X2A] average connectivity index chi-2
[X3A] average connectivity index chi-3
[X4A] average connectivity index chi-4
[X5A] average connectivity index chi-5
[X0v] valence connectivity index chi-0
[X1v] valence connectivity index chi-1
[X2v] valence connectivity index chi-2
[X3v] valence connectivity index chi-3
[X4v] valence connectivity index chi-4
[X5v] valence connectivity index chi-5
[X0Av] average valence connectivity index chi-0
[X1Av] average valence connectivity index chi-1
[X2Av] average valence connectivity index chi-2
[X3Av] average valence connectivity index chi-3
[X4Av] average valence connectivity index chi-4
[X5Av] average valence connectivity index chi-5

Balaban indices

Balaban Distance Connectivity Index, *J*

J is defined as (Balaban, 1982):

$$J = (B/(C+1))(\sum_{b \in B} (\sigma_i.\sigma_j)_b^{-1/2}),$$

where,

σ_i and σ_j are vertex distance degrees of two adjacent atoms *i* and *j* that are connected by the bond *b*, and

the sum runs over all the bonds *b* in the molecule,

B is the total number of bonds in the molecule, and

C is the cyclomatic number.

The **vertex distance degree** is the row sum σ_i of the Topological Level Matrix (see Appendix A), that is

$$\sigma_i = \sum_{j=1}^{n} d_{ij}$$

where,

d_{ij} is the ith entry in the Topological Level Matrix.

The **cyclomatic number** is the number of independent cycles *C* in a molecule or equivalently, the number of non-overlapping cycles. It is equal to the minimum number of edges that must be removed from the molecular graph to make it acyclic (that is, equal to the number of back edges in DFS tree). The cyclomatic number is given by:

$$C = B - n + 1,$$

where,

n is the number of atoms and *B* is the number of bonds in the molecule.

Balaban Modified Distance Connectivity Indices

Multi-graph distance matrix D^* is defined as a weighted distance matrix where the distance from an atom *i* to an atom *j* is obtained by counting the edges in the shortest path between them, where each edge *b* counts as the inverse of the conventional bond order π_b^*, that is,

$$[D^*]_{ij} = \sum_b 1/\pi_b^*,$$

where,

b represents the bond whose bond order is π_b^* and the sum runs through the sequence of bonds along the shortest path between the atoms *i* and *j*.

Conventional bond order π_b^* of a bond *b* is equal to 1, 2, 3 and 1.5 for single, double, triple and aromatic bonds respectively. Thus we can count each edge as 1, 0.5, 0.333 and 0.667 when it is single, double, triple and aromatic bonds respectively.

Multigraph distance degree σ_i^* of the ith atom is defined as the row sum in the Multigraph distance matrix, that is,

$$\sigma_i^* = \sum_{j=1}^{n} D_{ij}^*$$

The Balaban Modified Distance Connectivity Indices J^U, J^X, J^Y and J^Z are defined as follows (Balaban, 1986;

Balaban, 1990).

$$J^U = (B/(C+1))(\sum_{b \in B}(\sigma_i^* . \sigma_i^*)^{-1/2})$$

$$J^X = (B/(C+1))(\sum_{b \in B}(X_i . X_j . \sigma_i^* . \sigma_i^*)^{-1/2})$$

$$J^Y = (B/(C+1))(\sum_{b \in B}(Y_i . Y_j . \sigma_i^* . \sigma_i^*)^{-1/2})$$

$$J^V = (B/(C+1))(\sum_{b \in B}(V_i . V_j . \sigma_i^* . \sigma_i^*)^{-1/2})$$

where,

X_i, Y_i and V_i are electronegativity, covalent radius and van der Waal volume of the ith atom,

σ_i^* and σ_j^* are the multigraph distance degrees of two adjacent atoms *i* and *j* connected by the bond *b*, and

the sum runs over all the bonds *b* in the molecule,

B,is the total number of bonds in the molecule, and

C is the cyclomatic number.

Note that J^U is the unweighted Balaban Modified Distance Connectivity Index.

Balaban-type Indices

The **multigraph factor** for the ith atom is defined as:

$$f_i = \sum_{a \in V_i} \pi_{ia}^* - 1$$

where,

V_i is the set of neighbors of the ith atom, and

π_{ia}^* is the conventional bond order of the bond between the atoms *i* and *a*.

A **Balaban-type Index, DJ** is defined as follows (Balaban, 1993):

$$Jhetw = \sum_{i=1}^{n} \sum_{j \in V_i} (\mu_i . \mu_j)^{-1/2}$$

where,

$\mu_i = \sigma_i / \omega_i(1+f_i)$; σ_i is vertex distance degree and f_i is the multigraph factor for the ith atom (defined above),

V_i is the set of neighbors of the ith atom, that is the inner sum runs over all neighbors of the ith atom, and

w is one of *m, e, p, v, Z* which are relative mass, relative electronegativity, relative polarizability, relative van der waal volume and relative atomic number respectively, of the ith atom with respect to carbon.

The Balaban indices in sarchitect are:

[J] Balaban J index
[Uindex] Balaban U index
[Vindex] Balaban V index
[Xindex] Balaban X index
[Yindex] Balaban Y index
[Jhetm] Balaban-type index from mass weighted distance matrix
[Jhetv] Balaban-type index from van der Waals weighted distance matrix
[Jhete] Balaban-type index from electronegativity weighted distance matrix
[Jhetp] Balaban-type index from polarizability weighted distance matrix
[JhetZ] Balaban-type index from Z weighted distance matrix (Barysz matrix))

Charge indices

The Topological Charge Indices were proposed to evaluate the charge transfer between pairs of atoms and therefore, the global charge transfer in the molecule (Galvez, 1994; Galvez, 1995). The Topological Charge Indices are computed using the Topological Level Matrix, Reciprocal Square Distance Matrix and the Adjacency Matrix defined as follows.

The **Topological Level Matrix** contains the topological distance between a pair of atoms *i* and *j* at its ij^{th} entry.

The **Reciprocal Square Distance Matrix** D^{-2} is defined as the square matrix in which the ith entry is: $= 1/d_{ij}^2$ if $i \neq j$ and = 0 , otherwise,
where,

d_{ij} is the ij^{th} entry in the topological level matrix.

The **Adjacency Matrix** is defined as the square matrix in which the ij^{th} entry is:
=1 if *i* and *j* are connected by a bond = 0, otherwise.

The **Galvez Matrix** *M* is defined as:

$$M = A.D^{-2}.$$

here A and D^{-2} are as defined above.

The *vertex degree* δ_i of an atom i is defined as the number of atoms adjacent to it in the H-depleted molecular graph.

The **Charge Term Matrix** CT is defined as follows:

$$CT_{ij} = \delta_i \text{ if } i = j$$

where,

δ_i is the vertex degree of the ith atom. That is, the diagonal entries in the matrix are the vertex degrees of the corresponding atom, and $CT_{ij} = m_{ij} - m_{ji}$ if $i \neq j$

where,

m_{ij} and m_{ji} are the elements of the Galvez Matrix M. That is, the non-diagonal entries are the differences $m_{ij} - m_{ji}$.

The **Topological Charge Index** GG_k is defined as:

$$GG_k = 1/2 \sum_{i=1}^{n} \sum_{j=1}^{n} | CT_{ij} | \, \delta(k; d_{ij})$$

where,

n is the total number of non-Hydrogen atoms in the molecule,

$\delta(k; d_{ij})$ indicates Kronecker delta, namely:

$\delta(k; d_{ij}) = 1$, if $d_{ij} = k$ and

$\delta(k; d_{ij}) = 0$, otherwise.

The **Mean Topological Charge Index** JG_k is defined as:

$$JG_k = GG_k / (n - 1)$$

The **Global Topological Charge Index** (JGT) is given by:

$$JGT = \sum_{k=1}^{5} JG_k$$

[GGI1] topological charge index of order 1
[GGI2] topological charge index of order 2
[GGI3] topological charge index of order 3

[GGI4] topological charge index of order 4
[GGI5] topological charge index of order 5
[GGI6] topological charge index of order 6
[GGI7] topological charge index of order 7
[GGI8] topological charge index of order 8
[GGI9] topological charge index of order 9
[GGI10] topological charge index of order 10
[JGI1] mean topological charge index of order1
[JGI2] mean topological charge index of order2
[JGI3] mean topological charge index of order3
[JGI4] mean topological charge index of order4
[JGI5] mean topological charge index of order5
[JGI6] mean topological charge index of order6
[JGI7] mean topological charge index of order7
[JGI8] mean topological charge index of order8
[JGI9] mean topological charge index of order9
[JGI10] mean topological charge index of order10
[JGT] global topological charge index

Shape indices

Path/Walk Shape Indices

Atomic path/Walk Index for the ith atom is the ratio between Atomic Path Count P_i^m and Atomic walk Count $awc_i^{(m)}$ of same length *m, m* = 2, 3, 4,5 (Randic, 1999), i.e.

$$(p/w)_i^m = P_i^m / awc_i^{(m)}$$

Molecular Path/Walk Index is defined as the average sum of atomic path/walk indices of equal length:

$$(p/w)^m = (1/n).\sum_{i=1}^{n} (p/w)_i^m$$

The shape descriptor, **Petitjean Shape index** I_2 (labelled as PJI2) is a topological anisometry descriptor (Petitjean, 1992). It is defined as:

$$I_2 = (D - R) / R$$

[PW2] path/walk 2 - Randic shape index
[PW3] path/walk 3 - Randic shape index
[PW4] path/walk 4 - Randic shape index

[PW5] path/walk 5 - Randic shape index
[PJI2] Petiti Jean shape index

Eccentricity Indices

Let n be the total number of atoms (including Hydrogens) in a molecule. Consider the Topological Level Matrix (see Appendix A). For each atom i, the maximum value entry in the ith row is called the atom eccentricity v_i of the atom i. Based on this, the molecular graph can be characterized by two molecular descriptors, the topological radius $R = min_i (v_i)$ of the molecule and the topological diameter $D = max_i (v_i)$ of the molecule. Sarchitect has the following descriptors based on v_i, R and D (Konstantinova, 1996)

Eccentricity v (labelled as ECC):

$$v = \sum_{i=1}^{n} v_i$$

Average Atom Eccentricity $\bar{v}$ (labelled as AECC):

$$\bar{v} = v/n$$

Eccentric Δv (labelled as DECC):

$$\Delta v = (\sum_{i=1}^{n} (v_i - \bar{v})) / n$$

[ECC] eccentricity
[AECC] average eccentricity
[DECC] eccentric

Eigenvalue-based Indices

Eigenvalue Sum Descriptors

The Eigenvalue Sum Descriptors are computed from Weighted Distance Matrices of a Hydrogen-depleted Molecular Graph. The following weighted distance matrices are required for computation of the descriptors.

Barysz Distance Matrix D^Z: This is a weighted distance matrix accounting for the presence of heteroatoms and multiple bonds. The ij^{th} entry in the matrix D^Z is defined as follows (Barysz, 1983):

$$\{D^Z\}_{ij} = 1 - (Z_C / Z_i), \text{ if } i = j$$

$$\{D^Z\}_{ij} = \sum_{\delta=1}^{d_{ij}} Z_C^2 / (\pi_\delta^* . Z_{\delta(1)} Z_{\delta(2)}), \text{ if } i \neq j$$

where,

Z_C is the atomic number of carbon atom,

Z_i is the atomic number of the ij^{th} atom, and

the sum runs over all the bonds b in the shortest path from the atom i to the atom j, d_{ij} is the topological distance (i.e. the number of bonds from the atom i to the atom j in the molecule)

π_b^* is the conventional bond order of the bond b, that is 1,2,3 and 1.5 for single, double, triple and aromatic bonds respectively,

$Z_{b(1)}$ and $Z_{b(1)}$ are the atomic numbers of the atoms on the bond b

Electronegativity-weighted Distance Matrix D^X:

The Electronegativity-weighted Distance Matrix D^X is similar to the Barysz Distance Matrix. It can be obtained from the Barysz Distance Matrix by replacing atomic number by relative atom electronegativity (Ivanciuc, 1998). That is,

$$\{D^X\}_{ij} = 1-(X_C / X_i)$$, if $i \neq j$

$$\{D^X\}_{ij} = \sum_{b=1}^{d_{ij}} X_C^2 / (\pi_b^* . X_{b(1)} / X_{b(2)})$$, if $i \neq j$

where,

X_i is the relative atomic electronegativity of ij^{th} atom with respect to the carbon atom,

the sum runs over all the bonds b in the shortest path from the atom i to the atom j, d_{ij} is the topological distance.

π_b^* is the conventional bond order of the bond b, that is 1, 2, 3 and 1.5 for single, double, triple and aromatic bonds respectively,

$X_{b(1)}$ and $X_{b(2)}$ are the relative atomic electronegativity of the atoms forming the bond b

Covalent radius-weighted Distance Matrix D^Y:

The Covalent radius-weighted Distance Matrix D^Y is defined (Ivanciuc, 1998) as:

$$\{D^Y\}_{ij} = 1-(Y_C / Y_i)$$, if $i = j$

$$\{D^Y\}_{ij} = \sum_{b=1}^{d_{ij}} Y_C^2 / (\pi_b^* . Y_{b(1)} / Y_{b(2)})$$, if $i \neq j$

where,

Y_i is the relative atomic covalent radius of ith atom to the carbon atom.

Polarizability weighted distance matrix:

It is defined as:

$$\{D^P\}_{ij} = 1 - (P_C / P_i)$$, if $i = j$

$$\{D^P\}_{ij} = \sum_{\delta=1}^{d_{ij}} P_C^2 / (\pi_\delta^* . P_{\delta(1)} / P_{\delta(2)})$$, if $i \neq j$

where,

P_i is the relative atomic polarizability of ith atom to the carbon atom.

Mass weighted distance matrix:

It is defined as:

$$\{D^M\}_{ij} = 1 - (M_C / M_i)$$, if $i = j$

$$\{D^M\}_{ij} = \sum_{\delta=1}^{d_{ij}} M_C^2 / (\pi_\delta^* . M_{\delta(1)} / M_{\delta(2)})$$, if $i \neq j$

where,

M_i is the relative atomic mass of ith atom to the carbon atom

For each of the above defined matrices, the sum of their eigenvalues is computed to get the corresponding Eigenvalue Sum Descriptor.

VEA Indices

VEA indices (Balaban, 1991) are defined by the coefficients l_{iA} of the eigenvector associated with the largest negative eigenvalue (that is the last in the decreasing order of magnitude) of the Adjacency Matrix (see Appendix A).

We have the following three VEA Indices:

$$VEA1 = \sum_{i=1}^{A} l_{iA}$$

where A is the number of atoms in the molecule.

$VEA2 = (VEA1)/A$

$VEA3 = (A/10)\ (\log\ (VEA1))$

[SEigm] Eigenvalue sum from mass weighted distance matrix
[SEigv] Eigenvalue sum from van der Waals weighted distance matrix
[SEige] Eigenvalue sum from electronegativity weighted distance matrix
[SEigp] Eigenvalue sum from polarizability weighted distance matrix
[SEigZ] Eigenvalue sum from Z weighted distance matrix (Barysz matrix)
[VEA1] eigenvector coefficient sum from adjacency matrix
[VEA2] average eigenvector coefficient sum from adjacency matrix
[VEA3] log of eigenvector coefficient sum from adjacency matrix

E-state

The E-state Topological Parameter - TIE

The E-state Topological Parameter is a Balaban-type index derived from E-state indices and is similar to the Balaban Modified Distance Connectivity Indices (Voelkel, 1994). It is given as:

$$TIE = (B/(C+1))\left(\sum_{b \in B} (s_i s_j)_b^{-1/2}\right),$$

where,

s_i and s_j are the Electro-Topological State Indices for the two adjacent non-Hydrogen atoms i and j that are on the bond b, and

the sum runs over all the bonds b in the Hydrogen-depleted molecular graph,

B is the total number of bonds in the Hydrogen-depleted molecular graph, and

C is the cyclomatic number.

Aromaticity Indices

[BLI] Kier benzene-likeliness index

Super-Pendentic Index - SPI

Super-Pendentic Index is a topological descriptor computed from the H-depleted molecular graph (Gupta, 1999). By definition, it is calculated from the pendent matrix which is a sub-matrix of the topological distance matrix (see Appendix A) with one row for each atom and *m*, the number of columns corresponding to the number of terminal vertices (in the H-depleted molecular graph). The super-pendentic index is defined as:

$$\left(\sum_{i=1}^{n} \prod d_{ij}\right)^{1/2},$$

where,

n is the number of non-Hydrogen atoms, and

IId_{ij} is the product of the topological distances from the ith atom to all the terminal vertices *j*.

Topological Paths/Walks

Molecular path and walk counts

Walk and Path counts are atomic and molecular descriptors obtained from a H-depleted molecular graph based on the graph theory (Rucker & Rucker, 1993; Rucker & Rucker, 1994; Diudea, 1994; Randic, 1980; Razinger, 1986; Rucker & Rucker, 2000).

Molecular walk count

The adjacency matrix *A* of a molecular graph is an *n* by *n* matrix (where *n* is the number of atoms in the molecule) that has a value =1 at (*i, j*) if the atoms *i* and *j* are connected by a bond, and zero otherwise.

Let A^k be the kth power matrix of the adjacency matrix *A*, that is the product that we get when *A* is multiplied by itself *k* times. Its elements are denoted as $a_{ij}^{(k)}$.

Each entry (*i, j*) in the kth power matrix A^k denotes the number of walks of length *k* from the ith atom to j^{th} atom. The atomic walk count of order *k* for the ith atom denoted by $awc_i^{(k)}$ is given by the sum of the ith row elements in the kth power matrix A^k, given as:

$$awc_i^{(k)} = \sum_{j=1}^{n} a_{ij}^{(k)}$$

That is, $awc_i^{(k)}$ is the number of walks of length *k* originating from the ith atom.

Note that a walk can revisit a vertex any number of times.

Self-returning walk counts

A walk starting and ending on the same vertex, i.e. closed in itself is called a self-returning walk.

In particular, the diagonal elements (*i, j*) in the kth power matrix A^k denote the number of self-returning walks from the ith atom to itself

[MWC01] molecular walk count of order 1
[MWC02] molecular walk count of order 2
[MWC03] molecular walk count of order 3
[MWC04] molecular walk count of order 4
[MWC05] molecular walk count of order 5
[MWC06] molecular walk count of order 6
[MWC07] molecular walk count of order 7
[MWC08] molecular walk count of order 8
[MWC09] molecular walk count of order 9
[MWC10] molecular walk count of order 10
[TWC] total walk count
[SRW01] self-returning walk count of order 1
[SRW02] self-returning walk count of order 2
[SRW03] self-returning walk count of order 3
[SRW04] self-returning walk count of order 4
[SRW05] self-returning walk count of order 5
[SRW06] self-returning walk count of order 6
[SRW07] self-returning walk count of order 7
[SRW08] self-returning walk count of order 8
[SRW09] self-returning walk count of order 9
[SRW10] self-returning walk count of order 10

Molecular path count

A **Path** (or self-avoiding walk) is a walk with no vertex repeated. Atomic path count P_i^m is the number of paths of length *m* starting from the ith atom to any other atom in the molecule.

Molecular path count P^m is the count of all paths in the molecule of length *m*, that is,

$$P^m = \sum_{i=1}^{n} P_i^m$$

Molecular Multiple Path Count of Order m, piPCm is defined as follows. We define w_i, the weight of a path p_i as the sum of the bond orders of the bonds in the path p_i (the bond orders are respectively 1, 2, 3 and 1.5 for single, double, triple and aromatic bonds). piPCm is defined as the sum of the weights of the paths of length *m* in the molecule.

Meaning of Atomic Walk counts as a descriptor

Atomic Walk counts have the following significance: Atomic Walk count is a measure of something like "involvedness" or centrality of the atom in the molecular graph. It is a measure of the complexity of the vertex

environment. Atomic Walk count coincides with the extended connectivity defined by Morgan (Morgan, 1965; Razinger, 1982; Rucker and Rucker, 1993; Figueras, 1993)

[MPC01] molecular path count of order 1
[MPC02] molecular path count of order 2
[MPC03] molecular path count of order 3
[MPC04] molecular path count of order 4
[MPC05] molecular path count of order 5
[MPC06] molecular path count of order 6
[MPC07] molecular path count of order 7
[MPC08] molecular path count of order 8
[MPC09] molecular path count of order 9
[MPC10] molecular path count of order 10
[piPC01] molecular multiple path count of order 1
[piPC02] molecular multiple path count of order 2
[piPC03] molecular multiple path count of order 3
[piPC04] molecular multiple path count of order 4
[piPC05] molecular multiple path count of order 5
[piPC06] molecular multiple path count of order 6
[piPC07] molecular multiple path count of order 7
[piPC08] molecular multiple path count of order 8
[piPC09] molecular multiple path count of order 9
[piPC10] molecular multiple path count of order 10

Information Content Descriptors

Information Content Indices

The Information Content indices (Magnuson, 1983) are calculated based on the pairwise equivalence atoms in a Hydrogen-filled molecule. A pair of atoms are said to be equivalent at a particular level-r, if they are of the same element and their neighborhood is equivalent up to level-r. The equivalence of the neighborhoods of a pair of atoms *a* and *b* upto level-r is defined as follows: Let *set (a)* be the set of paths of length upto *r* originating from the atom *a*. Similarly, the *set (b)* corresponding to atom *b* is defined. Then the atoms *a* and *b* are said to be equivalent up to level-r, if there exists a one-to-one onto mapping between the sets *set (a)* and *set (b)*. We say that a mapping exists between a pair of paths if the corresponding atoms and bonds are the same in both the paths. For example, if the path C-C=N-O exists in each of the sets *set (a)* and *set (b)*, then these two paths can be mapped to each other.

The Information Content indices are calculated from level-0 to level-5 in Sarchitect, level zero.

In sarchitect the indices have been calculated from level-0 to level-5 and are given as:

Information Content Indices	

	[IC0] information content index (neighborhood symmetry of 0-order)
	[IC1] information content index (neighborhood symmetry of 1-order)
	[IC2] information content index (neighborhood symmetry of 2-order)
	[IC3] information content index (neighborhood symmetry of 3-order)
	[IC4] information content index (neighborhood symmetry of 4-order)
	[IC5] information content index (neighborhood symmetry of 5-order)
Bonding Information Content	
	[BIC0] bond information content (neighborhood symmetry of 0-order)
	[BIC1] bond information content (neighborhood symmetry of 1-order)
	[BIC2] bond information content (neighborhood symmetry of 2-order)
	[BIC3] bond information content (neighborhood symmetry of 3-order)
	[BIC4] bond information content (neighborhood symmetry of 4-order)
	[BIC5] bond information content (neighborhood symmetry of 5-order)
Structural Information Content	
	[SIC0] structural information content (neighborhood symmetry of 0-order)
	[SIC1] structural information content (neighborhood symmetry of 1-order)
	[SIC2] structural information content (neighborhood symmetry of 2-order)
	[SIC3] structural information content (neighborhood symmetry of 3-order)
	[SIC4] structural information content (neighborhood symmetry of 4-order)
	[SIC5] structural information content (neighborhood symmetry of 5-order)
Complementary Information Content	
	[CIC0] complementary information content (neighborhood symmetry of 0-order)
	[CIC1] complementary information content (neighborhood symmetry of 1-order)
	[CIC2] complementary information content (neighborhood symmetry of 2-order)
	[CIC3] complementary information content (neighborhood symmetry of 3-order)
	[CIC4] complementary information content (neighborhood symmetry of 4-order)
	[CIC5] complementary information content (neighborhood symmetry of 5-order)

Total Information Content	
	[TIC0] total information content index (neighborhood symmetry of 0-order)
	[TIC1] total information content index (neighborhood symmetry of 1-order)
	[TIC2] total information content index (neighborhood symmetry of 2-order)
	[TIC3] total information content index (neighborhood symmetry of 3-order)
	[TIC4] total information content index (neighborhood symmetry of 4-order)
	[TIC5] total information content index (neighborhood symmetry of 5-order)

Centric Indices

Lopping Centric Index - Lop

Lopping centric index (Balaban, 1979) is defined as the mean information content derived from the pruning partition of acyclic graphs. The pruning portions of an acyclic graph are obtained as follows. Vertices of degree one are removed from the graph recursively. Let A_i denote the number of vertices removed in the ith iteration and the graph will be empty after k iterations. Therefore the set of vertices in the graph is partitioned into $A_1, A_2, \ldots, A_k$, that is k partitions. Then the lopping centric index is defined as:

$$\sum_{i=1}^{k} (A_i / N) \log(A_i / N)$$

This definition is suitably modified in Sarchitect in order to include cyclic graphs.

Conformational (3D) Descriptors

Conformational descriptors are sensitive to the spatial positions of the atoms and their values vary for the same molecule depending upon the conformer chosen.

Randic profiles

Molecular profile

The Randic molecular profile DP_k (Randic 1995a, 1995b; Randic & Razinger 1995) is derived from the distance distribution moments of the geometric matrix G as the average row sum of its entries raised to the kth power and normalized by the factor $k!$.

$$DP_k = M_k / (k! \, n)$$

$$M_k = \sum_{i=1}^{n} \sum_{i=1}^{n} r_{ij}^{k}$$

where

r_{ij}^{k} is the kth power of the ij^{th} entry of the geometric matrix, and

n is the number of atoms.

As *k* increases, the contributions from the most distant pairs of atoms dominate. Sarchitect computes Randic molecular profiles for k=1,20:

DP01, DP02, DP03, DP04, DP05, DP06, DP07, DP08, DP09, DP11, DP12, DP13, DP14, DP15, DP16, DP17, DP18, DP19, DP20

Radial Distribution Function

The radial distribution function (RDF) descriptors are based on the distance distribution in the molecule. The radial distribution function of an ensemble of *n* atoms can be interpreted as the probability distribution of finding an atom in a spherical volume of radius *R*. A typical RDF descriptor is denoted by *RDFsw* where *s* and *w* take the values 10 ≤ *s* ≤ 155 in units of 5 and $w \in \{u, m, v, e, p\}$, and it is defined as follows:

$$RDF(R, w) = f \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_i w_j e^{-\beta(R - r_{ij})^2}$$

where,

f is a scaling factor,

r_{ij} is the Euclidean distance between the atoms *i* and *j*,

w_i and w_j are the weights of the atoms *i* and *j* respectively,

n is the total number of atoms,

β is the smoothing parameter which defines the probability distribution of the individual inter-atomic distance. *β* can be interpreted as the temperature factor that defines the movement of the atoms. We have chosen the value of 0.5 for *R* in our program.

The RDF descriptors given are:

Unweighted	
	[RDF010u] Radial Distribution Function - 1 / unweighted
	[RDF015u] Radial Distribution Function - 1.5 / unweighted
	[RDF020u] Radial Distribution Function - 2 / unweighted

	[RDF025u] Radial Distribution Function - 2.5 / unweighted
	[RDF030u] Radial Distribution Function - 3 / unweighted
	[RDF035u] Radial Distribution Function - 3.5 / unweighted
	[RDF040u] Radial Distribution Function - 4 / unweighted
	[RDF045u] Radial Distribution Function - 4.5 / unweighted
	[RDF050u] Radial Distribution Function - 5 / unweighted
	[RDF055u] Radial Distribution Function - 5.5 / unweighted
	[RDF060u] Radial Distribution Function - 6 / unweighted
	[RDF065u] Radial Distribution Function - 6.5 / unweighted
	[RDF070u] Radial Distribution Function - 7 / unweighted
	[RDF075u] Radial Distribution Function - 7.5 / unweighted
	[RDF080u] Radial Distribution Function - 8 / unweighted
	[RDF085u] Radial Distribution Function - 8.5 / unweighted
	[RDF090u] Radial Distribution Function - 9 / unweighted
	[RDF095u] Radial Distribution Function - 9.5 / unweighted
	[RDF100u] Radial Distribution Function - 10 / unweighted
	[RDF105u] Radial Distribution Function - 10.5 / unweighted
	[RDF110u] Radial Distribution Function - 11 / unweighted
	[RDF115u] Radial Distribution Function - 11.5 / unweighted
	[RDF120u] Radial Distribution Function - 12 / unweighted
	[RDF125u] Radial Distribution Function - 12.5 / unweighted
	[RDF130u] Radial Distribution Function - 13 / unweighted
	[RDF135u] Radial Distribution Function - 13.5 / unweighted
	[RDF140u] Radial Distribution Function - 14 / unweighted
	[RDF145u] Radial Distribution Function - 14.5 / unweighted
	[RDF150u] Radial Distribution Function - 15 / unweighted
	[RDF155u] Radial Distribution Function - 15.5 / unweighted
mass weighted	
	[RDF010m] Radial Distribution Function - 1 / weighted by atomic masses
	[RDF015m] Radial Distribution Function - 1.5 / weighted by atomic masses
	[RDF020m] Radial Distribution Function - 2 / weighted by atomic masses
	[RDF025m] Radial Distribution Function - 2.5 / weighted by atomic masses
	[RDF030m] Radial Distribution Function - 3 / weighted by atomic masses

	[RDF035m] Radial Distribution Function - 3.5 / weighted by atomic masses
	[RDF040m] Radial Distribution Function - 4 / weighted by atomic masses
	[RDF045m] Radial Distribution Function - 4.5 / weighted by atomic masses
	[RDF050m] Radial Distribution Function - 5 / weighted by atomic masses
	[RDF055m] Radial Distribution Function - 5.5 / weighted by atomic masses
	[RDF060m] Radial Distribution Function - 6 / weighted by atomic masses
	[RDF065m] Radial Distribution Function - 6.5 / weighted by atomic masses
	[RDF070m] Radial Distribution Function - 7 / weighted by atomic masses
	[RDF075m] Radial Distribution Function - 7.5 / weighted by atomic masses
	[RDF080m] Radial Distribution Function - 8 / weighted by atomic masses
	[RDF085m] Radial Distribution Function - 8.5 / weighted by atomic masses
	[RDF090m] Radial Distribution Function - 9 / weighted by atomic masses
	[RDF095m] Radial Distribution Function - 9.5 / weighted by atomic masses
	[RDF100m] Radial Distribution Function - 10 / weighted by atomic masses
	[RDF105m] Radial Distribution Function - 10.5 / weighted by atomic masses
	[RDF110m] Radial Distribution Function - 11 / weighted by atomic masses
	[RDF115m] Radial Distribution Function - 11.5 / weighted by atomic masses
	[RDF120m] Radial Distribution Function - 12 / weighted by atomic masses
	[RDF125m] Radial Distribution Function - 12.5 / weighted by atomic masses
	[RDF130m] Radial Distribution Function - 13 / weighted by atomic masses
	[RDF135m] Radial Distribution Function - 13.5 / weighted by atomic masses
	[RDF140m] Radial Distribution Function - 14 / weighted by atomic masses
	[RDF145m] Radial Distribution Function - 14.5 / weighted by atomic masses
	[RDF150m] Radial Distribution Function - 15 / weighted by atomic masses
	[RDF155m] Radial Distribution Function - 15.5 / weighted by atomic masses
volume weighted	
	[RDF010v] Radial Distribution Function - 1 / weighted by atomic van der Waals volumes
	[RDF015v] Radial Distribution Function - 1.5 / weighted by atomic van der Waals volumes
	[RDF020v] Radial Distribution Function - 2 / weighted by atomic van der Waals volumes
	[RDF025v] Radial Distribution Function - 2.5 / weighted by atomic van der Waals volumes
	[RDF030v] Radial Distribution Function - 3 / weighted by atomic van der Waals volumes
	[RDF035v] Radial Distribution Function - 3.5 / weighted by atomic van der Waals volumes

	[RDF040v] Radial Distribution Function - 4 / weighted by atomic van der Waals volumes
	[RDF045v] Radial Distribution Function - 4.5 / weighted by atomic van der Waals volumes
	[RDF050v] Radial Distribution Function - 5 / weighted by atomic van der Waals volumes
	[RDF055v] Radial Distribution Function - 5.5 / weighted by atomic van der Waals volumes
	[RDF060v] Radial Distribution Function - 6 / weighted by atomic van der Waals volumes
	[RDF065v] Radial Distribution Function - 6.5 / weighted by atomic van der Waals volumes
	[RDF070v] Radial Distribution Function - 7 / weighted by atomic van der Waals volumes
	[RDF075v] Radial Distribution Function - 7.5 / weighted by atomic van der Waals volumes
	[RDF080v] Radial Distribution Function - 8 / weighted by atomic van der Waals volumes
	[RDF085v] Radial Distribution Function - 8.5 / weighted by atomic van der Waals volumes
	[RDF090v] Radial Distribution Function - 9 / weighted by atomic van der Waals volumes
	[RDF095v] Radial Distribution Function - 9.5 / weighted by atomic van der Waals volumes
	[RDF100v] Radial Distribution Function - 10 / weighted by atomic van der Waals volumes
	[RDF105v] Radial Distribution Function - 10.5 / weighted by atomic van der Waals volumes
	[RDF110v] Radial Distribution Function - 11 / weighted by atomic van der Waals volumes
	[RDF115v] Radial Distribution Function - 11.5 / weighted by atomic van der Waals volumes
	[RDF120v] Radial Distribution Function - 12 / weighted by atomic van der Waals volumes
	[RDF125v] Radial Distribution Function - 12.5 / weighted by atomic van der Waals volumes
	[RDF130v] Radial Distribution Function - 13 / weighted by atomic van der Waals volumes
	[RDF135v] Radial Distribution Function - 13.5 / weighted by atomic van der Waals volumes
	[RDF140v] Radial Distribution Function - 14 / weighted by atomic van der Waals volumes
	[RDF145v] Radial Distribution Function - 14.5 / weighted by atomic van der Waals volumes
	[RDF150v] Radial Distribution Function - 15 / weighted by atomic van der Waals volumes
	[RDF155v] Radial Distribution Function - 15.5 / weighted by atomic van der Waals volumes
electronegativity weighted	
	[RDF010e] Radial Distribution Function - 1 / weighted by atomic Sanderson electronegativities
	[RDF015e] Radial Distribution Function - 1.5 / weighted by atomic Sanderson electronegativities
	[RDF020e] Radial Distribution Function - 2 / weighted by atomic Sanderson electronegativities

	[RDF025e] Radial Distribution Function - 2.5 / weighted by atomic Sanderson electronegativities
	[RDF030e] Radial Distribution Function - 3 / weighted by atomic Sanderson electronegativities
	[RDF035e] Radial Distribution Function - 3.5 / weighted by atomic Sanderson electronegativities
	[RDF040e] Radial Distribution Function - 4 / weighted by atomic Sanderson electronegativities
	[RDF045e] Radial Distribution Function - 4.5 / weighted by atomic Sanderson electronegativities
	[RDF050e] Radial Distribution Function - 5 / weighted by atomic Sanderson electronegativities
	[RDF055e] Radial Distribution Function - 5.5 / weighted by atomic Sanderson electronegativities
	[RDF060e] Radial Distribution Function - 6 / weighted by atomic Sanderson electronegativities
	[RDF065e] Radial Distribution Function - 6.5 / weighted by atomic Sanderson electronegativities
	[RDF070e] Radial Distribution Function - 7 / weighted by atomic Sanderson electronegativities
	[RDF075e] Radial Distribution Function - 7.5 / weighted by atomic Sanderson electronegativities
	[RDF080e] Radial Distribution Function - 8 / weighted by atomic Sanderson electronegativities
	[RDF085e] Radial Distribution Function - 8.5 / weighted by atomic Sanderson electronegativities
	[RDF090e] Radial Distribution Function - 9 / weighted by atomic Sanderson electronegativities
	[RDF095e] Radial Distribution Function - 9.5 / weighted by atomic Sanderson electronegativities
	[RDF100e] Radial Distribution Function - 10 / weighted by atomic Sanderson electronegativities
	[RDF105e] Radial Distribution Function - 10.5 / weighted by atomic Sanderson electronegativities
	[RDF110e] Radial Distribution Function - 11 / weighted by atomic Sanderson electronegativities
	[RDF115e] Radial Distribution Function - 11.5 / weighted by atomic Sanderson electronegativities
	[RDF120e] Radial Distribution Function - 12 / weighted by atomic Sanderson electronegativities
	[RDF125e] Radial Distribution Function - 12.5 / weighted by atomic Sanderson electronegativities
	[RDF130e] Radial Distribution Function - 13 / weighted by atomic Sanderson electronegativities
	[RDF135e] Radial Distribution Function - 13.5 / weighted by atomic Sanderson electronegativities

	[RDF140e] Radial Distribution Function - 14 / weighted by atomic Sanderson electronegativities
	[RDF145e] Radial Distribution Function - 14.5 / weighted by atomic Sanderson electronegativities
	[RDF150e] Radial Distribution Function - 15 / weighted by atomic Sanderson electronegativities
	[RDF155e] Radial Distribution Function - 15.5 / weighted by atomic Sanderson electronegativities
polarizability weighted	
	[RDF010p] Radial Distribution Function - 1 / weighted by atomic polarizabilities
	[RDF015p] Radial Distribution Function - 1.5 / weighted by atomic polarizabilities
	[RDF020p] Radial Distribution Function - 2 / weighted by atomic polarizabilities
	[RDF025p] Radial Distribution Function - 2.5 / weighted by atomic polarizabilities
	[RDF030p] Radial Distribution Function - 3 / weighted by atomic polarizabilities
	[RDF035p] Radial Distribution Function - 3.5 / weighted by atomic polarizabilities
	[RDF040p] Radial Distribution Function - 4 / weighted by atomic polarizabilities
	[RDF045p] Radial Distribution Function - 4.5 / weighted by atomic polarizabilities
	[RDF050p] Radial Distribution Function - 5 / weighted by atomic polarizabilities
	[RDF055p] Radial Distribution Function - 5.5 / weighted by atomic polarizabilities
	[RDF060p] Radial Distribution Function - 6 / weighted by atomic polarizabilities
	[RDF065p] Radial Distribution Function - 6.5 / weighted by atomic polarizabilities
	[RDF070p] Radial Distribution Function - 7 / weighted by atomic polarizabilities
	[RDF075p] Radial Distribution Function - 7.5 / weighted by atomic polarizabilities
	[RDF080p] Radial Distribution Function - 8 / weighted by atomic polarizabilities
	[RDF085p] Radial Distribution Function - 8.5 / weighted by atomic polarizabilities
	[RDF090p] Radial Distribution Function - 9 / weighted by atomic polarizabilities
	[RDF095p] Radial Distribution Function - 9.5 / weighted by atomic polarizabilities
	[RDF100p] Radial Distribution Function - 10 / weighted by atomic polarizabilities
	[RDF105p] Radial Distribution Function - 10.5 / weighted by atomic polarizabilities
	[RDF110p] Radial Distribution Function - 11 / weighted by atomic polarizabilities
	[RDF115p] Radial Distribution Function - 11.5 / weighted by atomic polarizabilities
	[RDF120p] Radial Distribution Function - 12 / weighted by atomic polarizabilities
	[RDF125p] Radial Distribution Function - 12.5 / weighted by atomic polarizabilities
	[RDF130p] Radial Distribution Function - 13 / weighted by atomic polarizabilities
	[RDF135p] Radial Distribution Function - 13.5 / weighted by atomic polarizabilities

	[RDF140p] Radial Distribution Function - 14 / weighted by atomic polarizabilities
	[RDF145p] Radial Distribution Function - 14.5 / weighted by atomic polarizabilities
	[RDF150p] Radial Distribution Function - 15 / weighted by atomic polarizabilities
	[RDF155p] Radial Distribution Function - 15.5 / weighted by atomic polarizabilities

MoRSE descriptors

3D MoRSE descriptors (3D Molecule Representation of Structures based on Electron diffraction) are derived from Infrared spectra simulation using a generalized scattering function (Soltzberg and Wilkins, 1977). A typical MoRSE descriptor is denoted by

M orsw where *s* and *w* take the values $1 \leq s \leq 32$ and $w \in \{u, m, v, e, p\}$, where,

- *u* is unweighted
- *m* is weighted by mass
- *v* is weighted by van der Waals volume
- *e* is weighted by electronegativity
- *p* is weighted by polarizability

The MoRSE descriptor is defined as follows:

$$Mor(s,w) = I(s,w) = \sum_{i=2}^{n} \sum_{j=1}^{i-1} w_i w_j \sin(s r_{ij}) / (s r_{ij}),$$

where,

r_{ij} is the Euclidean distance between the atoms *i* and *j* , and

w_i and w_j are the weights of the atoms *i* and *j* respectively.

(Schuur & Gasteiger, 1996; Schuur, 1996; Gasteiger, 1996; Schuur and Gasteiger, 1997)

The MoRSE descriptors given in sarchitect are:

Unweighted	
	[Mor01u] - signal 1 / unweighted
	[Mor02u] - signal 2 / unweighted
	[Mor03u] - signal 3 / unweighted
	[Mor04u] - signal 4 / unweighted
	[Mor05u] - signal 5 / unweighted

	[Mor06u] - signal 6 / unweighted
	[Mor07u] - signal 7 / unweighted
	[Mor08u] - signal 8 / unweighted
	[Mor09u] - signal 9 / unweighted
	[Mor10u] - signal 10 / unweighted
	[Mor11u] - signal 11 / unweighted
	[Mor12u] - signal 12 / unweighted
	[Mor13u] - signal 13 / unweighted
	[Mor14u] - signal 14 / unweighted
	[Mor15u] - signal 15 / unweighted
	[Mor16u] - signal 16 / unweighted
	[Mor17u] - signal 17 / unweighted
	[Mor18u] - signal 18 / unweighted
	[Mor19u] - signal 19 / unweighted
	[Mor20u] - signal 20 / unweighted
	[Mor21u] - signal 21 / unweighted
	[Mor22u] - signal 22 / unweighted
	[Mor23u] - signal 23 / unweighted
	[Mor24u] - signal 24 / unweighted
	[Mor25u] - signal 25 / unweighted
	[Mor26u] - signal 26 / unweighted
	[Mor27u] - signal 27 / unweighted
	[Mor28u] - signal 28 / unweighted
	[Mor29u] - signal 29 / unweighted
	[Mor30u] - signal 30 / unweighted
	[Mor31u] - signal 31 / unweighted
	[Mor32u] - signal 32 / unweighted
atomic mass weighted	
	[Mor01m] - signal 1 / weighted by atomic masses
	[Mor02m] - signal 2 / weighted by atomic masses
	[Mor03m] - signal 3 / weighted by atomic masses
	[Mor04m] - signal 4 / weighted by atomic masses
	[Mor05m] - signal 5 / weighted by atomic masses
	[Mor06m] - signal 6 / weighted by atomic masses
	[Mor07m] - signal 7 / weighted by atomic masses

	[Mor08m] - signal 8 / weighted by atomic masses
	[Mor09m] - signal 9 / weighted by atomic masses
	[Mor10m] - signal 10 / weighted by atomic masses
	[Mor11m] - signal 11 / weighted by atomic masses
	[Mor12m] - signal 12 / weighted by atomic masses
	[Mor13m] - signal 13 / weighted by atomic masses
	[Mor14m] - signal 14 / weighted by atomic masses
	[Mor15m] - signal 15 / weighted by atomic masses
	[Mor16m] - signal 16 / weighted by atomic masses
	[Mor17m] - signal 17 / weighted by atomic masses
	[Mor18m] - signal 18 / weighted by atomic masses
	[Mor19m] - signal 19 / weighted by atomic masses
	[Mor20m] - signal 20 / weighted by atomic masses
	[Mor21m] - signal 21 / weighted by atomic masses
	[Mor22m] - signal 22 / weighted by atomic masses
	[Mor23m] - signal 23 / weighted by atomic masses
	[Mor24m] - signal 24 / weighted by atomic masses
	[Mor25m] - signal 25 / weighted by atomic masses
	[Mor26m] - signal 26 / weighted by atomic masses
	[Mor27m] - signal 27 / weighted by atomic masses
	[Mor28m] - signal 28 / weighted by atomic masses
	[Mor29m] - signal 29 / weighted by atomic masses
	[Mor30m] - signal 30 / weighted by atomic masses
	[Mor31m] - signal 31 / weighted by atomic masses
	[Mor32m] - signal 32 / weighted by atomic masses
vdw volume weighted	
	[Mor01v] - signal 1 / weighted by atomic van der Waals volumes
	[Mor02v] - signal 2 / weighted by atomic van der Waals volumes
	[Mor03v] - signal 3 / weighted by atomic van der Waals volumes
	[Mor04v] - signal 4 / weighted by atomic van der Waals volumes
	[Mor05v] - signal 5 / weighted by atomic van der Waals volumes
	[Mor06v] - signal 6 / weighted by atomic van der Waals volumes
	[Mor07v] - signal 7 / weighted by atomic van der Waals volumes
	[Mor08v] - signal 8 / weighted by atomic van der Waals volumes
	[Mor09v] - signal 9 / weighted by atomic van der Waals volumes

	[Mor10v] - signal 10 / weighted by atomic van der Waals volumes
	[Mor11v] - signal 11 / weighted by atomic van der Waals volumes
	[Mor12v] - signal 12 / weighted by atomic van der Waals volumes
	[Mor13v] - signal 13 / weighted by atomic van der Waals volumes
	[Mor14v] - signal 14 / weighted by atomic van der Waals volumes
	[Mor15v] - signal 15 / weighted by atomic van der Waals volumes
	[Mor16v] - signal 16 / weighted by atomic van der Waals volumes
	[Mor17v] - signal 17 / weighted by atomic van der Waals volumes
	[Mor18v] - signal 18 / weighted by atomic van der Waals volumes
	[Mor19v] - signal 19 / weighted by atomic van der Waals volumes
	[Mor20v] - signal 20 / weighted by atomic van der Waals volumes
	[Mor21v] - signal 21 / weighted by atomic van der Waals volumes
	[Mor22v] - signal 22 / weighted by atomic van der Waals volumes
	[Mor23v] - signal 23 / weighted by atomic van der Waals volumes
	[Mor24v] - signal 24 / weighted by atomic van der Waals volumes
	[Mor25v] - signal 25 / weighted by atomic van der Waals volumes
	[Mor26v] - signal 26 / weighted by atomic van der Waals volumes
	[Mor27v] - signal 27 / weighted by atomic van der Waals volumes
	[Mor28v] - signal 28 / weighted by atomic van der Waals volumes
	[Mor29v] - signal 29 / weighted by atomic van der Waals volumes
	[Mor30v] - signal 30 / weighted by atomic van der Waals volumes
	[Mor31v] - signal 31 / weighted by atomic van der Waals volumes
	[Mor32v] - signal 32 / weighted by atomic van der Waals volumes
electronegativity weighted	
	[Mor01e] - signal 1 / weighted by atomic Sanderson electronegativities
	[Mor02e] - signal 2 / weighted by atomic Sanderson electronegativities
	[Mor03e] - signal 3 / weighted by atomic Sanderson electronegativities
	[Mor04e] - signal 4 / weighted by atomic Sanderson electronegativities
	[Mor05e] - signal 5 / weighted by atomic Sanderson electronegativities
	[Mor06e] - signal 6 / weighted by atomic Sanderson electronegativities
	[Mor07e] - signal 7 / weighted by atomic Sanderson electronegativities
	[Mor08e] - signal 8 / weighted by atomic Sanderson electronegativities
	[Mor09e] - signal 9 / weighted by atomic Sanderson electronegativities
	[Mor10e] - signal 10 / weighted by atomic Sanderson electronegativities

	[Mor11e] - signal 11 / weighted by atomic Sanderson electronegativities
	[Mor12e] - signal 12 / weighted by atomic Sanderson electronegativities
	[Mor13e] - signal 13 / weighted by atomic Sanderson electronegativities
	[Mor14e] - signal 14 / weighted by atomic Sanderson electronegativities
	[Mor15e] - signal 15 / weighted by atomic Sanderson electronegativities
	[Mor16e] - signal 16 / weighted by atomic Sanderson electronegativities
	[Mor17e] - signal 17 / weighted by atomic Sanderson electronegativities
	[Mor18e] - signal 18 / weighted by atomic Sanderson electronegativities
	[Mor19e] - signal 19 / weighted by atomic Sanderson electronegativities
	[Mor20e] - signal 20 / weighted by atomic Sanderson electronegativities
	[Mor21e] - signal 21 / weighted by atomic Sanderson electronegativities
	[Mor22e] - signal 22 / weighted by atomic Sanderson electronegativities
	[Mor23e] - signal 23 / weighted by atomic Sanderson electronegativities
	[Mor24e] - signal 24 / weighted by atomic Sanderson electronegativities
	[Mor25e] - signal 25 / weighted by atomic Sanderson electronegativities
	[Mor26e] - signal 26 / weighted by atomic Sanderson electronegativities
	[Mor27e] - signal 27 / weighted by atomic Sanderson electronegativities
	[Mor28e] - signal 28 / weighted by atomic Sanderson electronegativities
	[Mor29e] - signal 29 / weighted by atomic Sanderson electronegativities
	[Mor30e] - signal 30 / weighted by atomic Sanderson electronegativities
	[Mor31e] - signal 31 / weighted by atomic Sanderson electronegativities
	[Mor32e] - signal 32 / weighted by atomic Sanderson electronegativities
polarizability weighted	
	[Mor01p] - signal 1 / weighted by atomic polarizabilities
	[Mor02p] - signal 2 / weighted by atomic polarizabilities
	[Mor03p] - signal 3 / weighted by atomic polarizabilities
	[Mor04p] - signal 4 / weighted by atomic polarizabilities
	[Mor05p] - signal 5 / weighted by atomic polarizabilities
	[Mor06p] - signal 6 / weighted by atomic polarizabilities
	[Mor07p] - signal 7 / weighted by atomic polarizabilities
	[Mor08p] - signal 8 / weighted by atomic polarizabilities
	[Mor09p] - signal 9 / weighted by atomic polarizabilities
	[Mor10p] - signal 10 / weighted by atomic polarizabilities
	[Mor11p] - signal 11 / weighted by atomic polarizabilities
	[Mor12p] - signal 12 / weighted by atomic polarizabilities

	[Mor13p] - signal 13 / weighted by atomic polarizabilities
	[Mor14p] - signal 14 / weighted by atomic polarizabilities
	[Mor15p] - signal 15 / weighted by atomic polarizabilities
	[Mor16p] - signal 16 / weighted by atomic polarizabilities
	[Mor17p] - signal 17 / weighted by atomic polarizabilities
	[Mor18p] - signal 18 / weighted by atomic polarizabilities
	[Mor19p] - signal 19 / weighted by atomic polarizabilities
	[Mor20p] - signal 20 / weighted by atomic polarizabilities
	[Mor21p] - signal 21 / weighted by atomic polarizabilities
	[Mor22p] - signal 22 / weighted by atomic polarizabilities
	[Mor23p] - signal 23 / weighted by atomic polarizabilities
	[Mor24p] - signal 24 / weighted by atomic polarizabilities
	[Mor25p] - signal 25 / weighted by atomic polarizabilities
	[Mor26p] - signal 26 / weighted by atomic polarizabilities
	[Mor27p] - signal 27 / weighted by atomic polarizabilities
	[Mor28p] - signal 28 / weighted by atomic polarizabilities
	[Mor29p] - signal 29 / weighted by atomic polarizabilities
	[Mor30p] - signal 30 / weighted by atomic polarizabilities
	[Mor31p] - signal 31 / weighted by atomic polarizabilities
	[Mor32p] - signal 32 / weighted by atomic polarizabilities

WHIM descriptors

WHIM descriptors are based on the statistical indices calculated on the projections of atoms along principal axes (Todeschini, 1994; Todeschini and Gramatica, 1997a). They are built in such a way as to capture relevant molecular 3D information regarding the molecular size, shape, symmetry and atom distribution with respect to invariant reference frames. The algorithm consists of performing a Principal Components Analysis on the centered Cartesian coordinates of a molecule by using a weighted covariance matrix obtained from different weighing schemes for the atoms. The following weighting schemes are used for computing the weighted covariance matrix, S^w:

- unweighted (u), that is the weight $w_i = 1$ for each i.
- atomic masses ($w_i = m_i$)
- atomic van der Waals volumes ($w_i = v_i$)
- atomic Sanderson electronegativities ($w_i = e_i$)
- atomic polarizabilities ($w_i = p_i$)
- atomic electro topological states ($w_i = s_i$) (See 'Computing the Electro topological State Indices' at the end of this section for more details).

The weighted covariance matrix is a 3×3 matrix whose elements are the weighted covariances S_{jk} between j^{th} and the kth atomic coordinates for $j,k \in \{1,2,3\}$ defined as follows:

$$s_{jk}^{w} = \sum_{i=1}^{n} w_i (q_{ij} - \bar{q}_j)(q_{ik} - \bar{q}_k) / \sum_{i=1}^{n} w_i$$

where, n is the number of atoms, w_i is the weight of the ith atom, q_{ij} and q_{ik} are the jth and the kth coordinates of the ith atom and $\bar{q}_j = 1/n\sum_{i=1}^{n} q_{ij}$ and $\bar{q}_k = 1/n\sum_{i=1}^{n} q_{ik}$ w_i the weight of the ith atom is in one of the six weighting schemes defined above.

Note that each of these weights are scaled with respect to the carbon atom. To make all the weights positive, the electro-topological indices are scaled thus:

$$S_i^d = S_i + 7$$

Further, when ever we compute WHIM descriptors with the electro-topological states, we consider only the non-Hydrogen atoms.

Depending on the kind of weighting scheme, different covariance matrices and different principal axes (principal components t_m) are obtained. For example, using atomic masses as the weighting scheme, the directions of the three principal axes are the directions of the principal inertia axes. Thus, the WHIM approach can be viewed as a generalization searching for the principal axes with respect to a defined atomic property (the weighting scheme). For each weighting scheme, a set of statistical indices is calculated on the atoms projected onto each principal component t_m m = 1, 2, 3 i.e. the scores.

Directional WHIM Descriptors

These are molecular descriptors calculated as univariate statistical indices on the scores of each individual principal component t_m m = 1, 2, 3.

Directional WHIM Descriptors: WHIM size: Directional WHIM size descriptors are the eigenvalues λ_1, λ_2 and λ_3 of the matrix S^w, named as: $L1w = \lambda_1$, $L2w = \lambda_2$, $L3w = \lambda_3$.

Directional WHIM Descriptors: WHIM shape: Directional WHIM shape descriptors are v_1, v_2, v_3 defined as:

$$Piw = v_i = \lambda_i /(\lambda_1 + \lambda_2 + \lambda_3)$$

For i = 1, 2, 3. Since $v_1 + v_2 + v_3 = 1$, only two of them are important.

Directional WHIM Descriptors: WHIM symmetry: WHIM symmetry indices γ_i , i = 1, 2, 3, are defined as follows:

$$\gamma_i = -[(n_{si}/n)\log_2(n_{si}/n) + n_{ai}((1/n)\log_2(1/n))]$$

for i = 1, 2, 3 and

$$Giw = \gamma_i = 1/(1+\gamma_i),$$

for i = 1, 2, 3, where, n_{si} is the number of central symmetric atoms, n_{ai} is the number of central asymmetric atoms (along the ith component), and n is the total number of atoms.

Directional WHIM Descriptors: WHIM accessibility: WHIM accessibility descriptor *Emw* is computed from Kurtosis.

$$k_m = (\sum_{i=1}^{n} t_{mi}^4) / \lambda_m^2 . n$$

for m = 1, 2, 3, where,

$$t_{mi} = \sum_{j \in \{x,y,z\}} (q_{ij} - \bar{q}_j)(\lambda_m(j)),$$

and λ_m (j) is the jth component of the vector λ_m .

the WHIM accessibility descriptor is given by,

$$Emw = 1/ k_m$$

for ***m* = 1, 2, 3**,.

Global WHIM Descriptors:

Global WHIM Descriptors are calculated directly as a combination of the directional WHIM descriptors.

Global WHIM Descriptors: Total size-1:

$$Tw = L1w + L2w + L3w$$

Global WHIM Descriptors: Total size-2:

$$Aw = (L1w. L2w) + (L2w. L3w) + (L3w. L1w)$$

Global WHIM Descriptors: Total size-3:

$$Vw = Tw + Aw + (L1w + L2w + L3w)$$

Global WHIM Descriptors: Total symmetry:

$$Gm = (G1w. G2w. G3w)^{1/3}$$

Global WHIM Descriptors: Total accessibility:

$$Kw = 3/4.\sum_i |(Liw/(L1w + L2w + L3w) - 1/3|$$

Global WHIM Descriptors: Total density:

$$Dw = E1w + E2w + E3w$$

Computing the Electro topological State Indices

The Electro topological State S_i of the ith atom in a molecule, also called the E-state index gives information related to the electronic and topological state of the atom in the molecule defined as:

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

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 the ith atom by all other atoms in the molecule, d_{ij} is the topological distance between the atoms i and j, that is the ij^{th} entry in the Topological Level Matrix, and n is the number of non-Hydrogen atoms. The exponent k is the parameter to modify the influence of distant or the nearby atoms for particular studies. The recommended value for k = 2.

The intrinsic state of the ith atom is calculated by:

$$I_i = ((2/L_i)^2 \delta_i^v + 1)/\delta_i,$$

where, L_i is the principal quantum number, δ_i^v is the valence vertex degree, and δ_i is the vertex degree of the ith atom in the Hydrogen-depleted molecular graph. The vertex degree of an atom is the number of non-Hydrogen neighbors of that atom, whereas the valence vertex degree is defined as follows:

If the atom is of second principal quantum level (C, N, O, F) then its valence vertex degree is defined as:

$$\delta_i^v = Z_i^v - h_i = \sigma_i + \pi_i + n_i - h_i,$$

where, Z_i^v is the number of valence electrons (that is, σ electrons, π electrons and lone pair electrons n) of the ith atom, and h_i is the number of Hydrogens bonded to that atom.

If the atom is of higher than second principal quantum level (P, S, Cl, Br, I) then its valence vertex degree is defined as:

$$\delta_i^v = (Z_i^v - h_i)/(Z_i - Z_i^v - 1),$$

where, Z_i^v is the number of valence electrons (that is, σ electrons, π electrons and lone pair electrons n) of the ith atom, h_i is the number of Hydrogens bonded to that atom, and Z_i is the total number of electrons in the ith atom, that is its atomic number.

mass weighted	
	[L1m] 1st component size directional WHIM index / weighted by atomic masses
	[L2m] 2nd component size directional WHIM index / weighted by atomic masses
	[L3m] 3rd component size directional WHIM index / weighted by atomic masses
	[P1m] 1st component shape directional WHIM index / weighted by atomic masses
	[P2m] 2nd component shape directional WHIM index / weighted by atomic masses
	[E1m] 1st component accessibility directional WHIM index / weighted by atomic masses
	[E2m] 2nd component accessibility directional WHIM index / weighted by atomic masses
	[E3m] 3rd component accessibility directional WHIM index / weighted by atomic masses
	[Tm] T total size index / weighted by atomic masses
	[Am] A total size index / weighted by atomic masses
	[Vm] V total size index / weighted by atomic masses
	[Km] K global shape index / weighted by atomic masses
	[Dm] D total accessibility index / weighted by atomic masses
vdw volume weighted	[L1v] 1st component size directional WHIM index / weighted by atomic van der Waals volumes
	[L2v] 2nd component size directional WHIM index / weighted by atomic van der Waals volumes
	[L3v] 3rd component size directional WHIM index / weighted by atomic van der Waals volumes
	[P1v] 1st component shape directional WHIM index / weighted by atomic van der Waals volumes
	[P2v] 2nd component shape directional WHIM index / weighted by atomic van der Waals volumes
	[E1v] 1st component accessibility directional WHIM index / weighted by atomic van der Waals volumes
	[E2v] 2nd component accessibility directional WHIM index / weighted by atomic van der Waals volumes
	[E3v] 3rd component accessibility directional WHIM index / weighted by atomic van der Waals volumes
	[Tv] T total size index / weighted by atomic van der Waals volumes
	[Av] A total size index / weighted by atomic van der Waals volumes
	[Vv] V total size index / weighted by atomic van der Waals volumes

	[Kv] K global shape index / weighted by atomic van der Waals volumes
	[Dv] D total accessibility index / weighted by atomic van der Waals volumes
electronegativity weighted	[L1e] 1st component size directional WHIM index / weighted by atomic Sanderson electronegativities
	[L2e] 2nd component size directional WHIM index / weighted by atomic Sanderson electronegativities
	[L3e] 3rd component size directional WHIM index / weighted by atomic Sanderson electronegativities
	[P1e] 1st component shape directional WHIM index / weighted by atomic Sanderson electronegativities
	[P2e] 2nd component shape directional WHIM index / weighted by atomic Sanderson electronegativities
	[E1e] 1st component accessibility directional WHIM index / weighted by atomic Sanderson electronegativities
	[E2e] 2nd component accessibility directional WHIM index / weighted by atomic Sanderson electronegativities
	[E3e] 3rd component accessibility directional WHIM index / weighted by atomic Sanderson electronegativities
	[Te] T total size index / weighted by atomic Sanderson electronegativities
	[Ae] A total size index / weighted by atomic Sanderson electronegativities
	[Ve] V total size index / weighted by atomic Sanderson electronegativities
	[Ke] K global shape index / weighted by atomic Sanderson electronegativities
	[De] D total accessibility index / weighted by atomic Sanderson electronegativities
polarizability weighted	
	[L1p] 1st component size directional WHIM index / weighted by atomic polarizabilities
	[L2p] 2nd component size directional WHIM index / weighted by atomic polarizabilities
	[L3p] 3rd component size directional WHIM index / weighted by atomic polarizabilities
	[P1p] 1st component shape directional WHIM index / weighted by atomic polarizabilities
	[P2p] 2nd component shape directional WHIM index / weighted by atomic polarizabilities
	[E1p] 1st component accessibility directional WHIM index / weighted by atomic polarizabilities
	[E2p] 2nd component accessibility directional WHIM index / weighted by atomic polarizabilities
	[E3p] 3rd component accessibility directional WHIM index / weighted by atomic polarizabilities
	[Tp] T total size index / weighted by atomic polarizabilities
	[Ap] A total size index / weighted by atomic polarizabilities

	[Vp] V total size index / weighted by atomic polarizabilities
	[Kp] K global shape index / weighted by atomic polarizabilities
	[Dp] D total accessibility index / weighted by atomic polarizabilities
Unweighted	
	[L1u] 1st component size directional WHIM index / unweighted
	[L2u] 2nd component size directional WHIM index / unweighted
	[L3u] 3rd component size directional WHIM index / unweighted
	[P1u] 1st component shape directional WHIM index / unweighted
	[P2u] 2nd component shape directional WHIM index / unweighted
	[E1u] 1st component accessibility directional WHIM index / unweighted
	[E2u] 2nd component accessibility directional WHIM index / unweighted
	[E3u] 3rd component accessibility directional WHIM index / unweighted
	[Tu] T total size index / unweighted
	[Au] A total size index / unweighted
	[Vu] V total size index / unweighted
	[Ku] K global shape index / unweighted
	[Du] D total accessibility index / unweighted
electrotopological state weighted	
	[L1s] 1st component size directional WHIM index / weighted by atomic electrotopological states
	[L2s] 2nd component size directional WHIM index / weighted by atomic electrotopological states
	[L3s] 3rd component size directional WHIM index / weighted by atomic electrotopological states
	[P1s] 1st component shape directional WHIM index / weighted by atomic electrotopological states
	[P2s] 2nd component shape directional WHIM index / weighted by atomic electrotopological states
	[E1s] 1st component accessibility directional WHIM index / weighted by atomic electrotopological states
	[E2s] 2nd component accessibility directional WHIM index / weighted by atomic electrotopological states
	[E3s] 3rd component accessibility directional WHIM index / weighted by atomic electrotopological states
	[Ts] T total size index / weighted by atomic electrotopological states
	[As] A total size index / weighted by atomic electrotopological states
	[Vs] V total size index / weighted by atomic electrotopological states
	[Ks] K global shape index / weighted by atomic electrotopological states

	[Ds] D total accessibility index / weighted by atomic electrotopological states

GETAWAY descriptors

The GETAWAY (GEometry, Topology, and Atom-Weights AssemblY) descriptors are molecular descriptors derived from the Molecular Influence Matrix (MIM). The following four types of GETAWAY descriptors are computed namely *HATSkw*, *Hkw*, *Rkw* and $R^{+}kw$, where $1 \leq k \leq 8$, and $w \in \{u, m, e, v, p,\}$ as in the Autocorrelation Descriptors except for the additional type *u* defined as unweighted, where $w_i = 1$ for all atoms *i*.

HATS descriptors

HATS descriptors are computed on a Hydrogen-filled molecule. We construct the Molecular Influence matrix *H* as follows. Let *M* be the geometric distance matrix having *n* rows and 3 columns, where we have one row for each of the *n* atoms present in the molecule and one column for each of the x-, y-, z- coordinates of the atoms in the molecule. The atomic coordinates are assumed to be calculated with respect to the geometric centre of the molecule. The Molecular Influence matrix *H* is obtained from *M* as follows:

$$H = M.\ (M^T.M)^{-1}.M^T,$$

where, M^T is the transpose of *M* and $(M^T.M)^{-1}$ is the inverse of $(M^T.M)$.

Each of the HATS descriptors is of the form *HATSkw* where $1 \leq k \leq 8$ and $w \in \{u, m, e, v, p,\}$ as in the Autocorrelation Descriptors except for the additional type *u* defined as unweighted, where $w_i = 1$ for all atoms *i*.

HATSkw is defined as follows. Let the atoms *i* and *j* be separated by a lag *k* as in other autocorrelation descriptors (say ATS). Let h_{ii} and h_{jj} be the diagonal entries corresponding to the atoms *i* and *j* in the *H* Matrix. Then we define HATSkw as follows:

$$HATSkw = \sum_{i=1}^{n}\sum_{j=1}^{n}(w_i h_{ii}).(w_j h_{jj})\,\delta(k; d_{ij}),$$

$\delta(k; D_{ij})$ is Kronecker delta, that is, $\delta(k; d_{ij}) = 1$ if the ij^{th} entry in the Topological Level Matrix is $= k$, and $\delta(k; d_{ij}) = 1$ otherwise.

Hkw

Hkw is defined as:

$$Hkw = \sum_{i=1}^{n}\sum_{j=1}^{n}(h_{ij}.w_i.w_j).\delta(k; d_{ij}; h_{ij}),$$

where, $\delta(k; d_{ij}; h_{ij}) = 1$ if $d_{ij} = k$ (in the Topological Level Matrix) AND $h_{ij} > 0$ (in the *H* matrix), and zero otherwise.

HTw

HTw is defined as:

$$HTw = H0w + 2\sum_{k=1}^{8} Hkw$$

Rkw

We define the Influence/Distance Matrix R as follows. The diagonal elements R (*ii*) = 0 for each *i*. R(*ij*) $= (\sqrt{h_{ii}.h_{jj}}) / r_{ij}$ for $i \neq j$, where r_{ij} is the geometric distance between the atoms *i* and *j*.

Rkw is defined as:

$$Rkw = \sum_{i=1}^{n-1}\sum_{j>1}^{n} ((\sqrt{h_{ii}.h_{jj}}) / r_{ij}).w_i.w_j.\delta(k, d_{ij}),$$

where, $\delta(k; d_{ij}) = 1$ if $d_{ij} = k$ (in the Topological Level Matrix), and zero otherwise.

The highest value of $((\sqrt{h_{ii}.h_{jj}}) / r_{ij}).w_i.w_j.\delta(k, d_{ij})$ for any particular lag k and particular *w* is denoted by the R_k^+w, that is,

$$R^+kw = \max_{ij}(((\sqrt{h_{ii}.h_{jj}}) / r_{ij}).w_i.w_j.\delta(k, d_{ij}))$$

mass weighted	
	[HATS0m] leverage-weighted autocorrelation of lag 0 / weighted by atomic masses
	[HATS1m] leverage-weighted autocorrelation of lag 1 / weighted by atomic masses
	[HATS2m] leverage-weighted autocorrelation of lag 2 / weighted by atomic masses
	[HATS3m] leverage-weighted autocorrelation of lag 3 / weighted by atomic masses
	[HATS4m] leverage-weighted autocorrelation of lag 4 / weighted by atomic masses
	[HATS5m] leverage-weighted autocorrelation of lag 5 / weighted by atomic masses
	[HATS6m] leverage-weighted autocorrelation of lag 6 / weighted by atomic masses
	[HATS7m] leverage-weighted autocorrelation of lag 7 / weighted by atomic masses
	[HATS8m] leverage-weighted autocorrelation of lag 8 / weighted by atomic masses
	[H0m] H autocorrelation of lag 0 / weighted by atomic masses
	[H1m] H autocorrelation of lag 1 / weighted by atomic masses
	[H2m] H autocorrelation of lag 2 / weighted by atomic masses
	[H3m] H autocorrelation of lag 3 / weighted by atomic masses
	[H4m] H autocorrelation of lag 4 / weighted by atomic masses
	[H5m] H autocorrelation of lag 5 / weighted by atomic masses
	[H6m] H autocorrelation of lag 6 / weighted by atomic masses
	[H7m] H autocorrelation of lag 7 / weighted by atomic masses
	[H8m] H autocorrelation of lag 8 / weighted by atomic masses
	[R1m] R autocorrelation of lag 1 / weighted by atomic masses
	[R2m] R autocorrelation of lag 2 / weighted by atomic masses
	[R3m] R autocorrelation of lag 3 / weighted by atomic masses
	[R4m] R autocorrelation of lag 4 / weighted by atomic masses
	[R5m] R autocorrelation of lag 5 / weighted by atomic masses
	[R6m] R autocorrelation of lag 6 / weighted by atomic masses
	[R7m] R autocorrelation of lag 7 / weighted by atomic masses
	[R8m] R autocorrelation of lag 8 / weighted by atomic masses
	[R1m+] R maximal autocorrelation of lag 1 / weighted by atomic masses
	[R2m+] R maximal autocorrelation of lag 2 / weighted by atomic masses
	[R3m+] R maximal autocorrelation of lag 3 / weighted by atomic masses
	[R4m+] R maximal autocorrelation of lag 4 / weighted by atomic masses

	[R5m+] R maximal autocorrelation of lag 5 / weighted by atomic masses
	[R6m+] R maximal autocorrelation of lag 6 / weighted by atomic masses
	[R7m+] R maximal autocorrelation of lag 7 / weighted by atomic masses
	[R8m+] R maximal autocorrelation of lag 8 / weighted by atomic masses
	[HATSm] leverage-weighted total index / weighted by atomic masses
	[HTm] H total index / weighted by atomic masses
	[RTm] R total index / weighted by atomic masses
vdw volume weighted	
	[HATS0v] leverage-weighted autocorrelation of lag 0 / weighted by atomic van der Waals volumes
	[HATS1v] leverage-weighted autocorrelation of lag 1 / weighted by atomic van der Waals volumes
	[HATS2v] leverage-weighted autocorrelation of lag 2 / weighted by atomic van der Waals volumes
	[HATS3v] leverage-weighted autocorrelation of lag 3 / weighted by atomic van der Waals volumes
	[HATS4v] leverage-weighted autocorrelation of lag 4 / weighted by atomic van der Waals volumes
	[HATS5v] leverage-weighted autocorrelation of lag 5 / weighted by atomic van der Waals volumes
	[HATS6v] leverage-weighted autocorrelation of lag 6 / weighted by atomic van der Waals volumes
	[HATS7v] leverage-weighted autocorrelation of lag 7 / weighted by atomic van der Waals volumes
	[HATS8v] leverage-weighted autocorrelation of lag 8 / weighted by atomic van der Waals volumes
	[H0v] H autocorrelation of lag 0 / weighted by atomic van der Waals volumes
	[H1v] H autocorrelation of lag 1 / weighted by atomic van der Waals volumes
	[H2v] H autocorrelation of lag 2 / weighted by atomic van der Waals volumes
	[H3v] H autocorrelation of lag 3 / weighted by atomic van der Waals volumes
	[H4v] H autocorrelation of lag 4 / weighted by atomic van der Waals volumes
	[H5v] H autocorrelation of lag 5 / weighted by atomic van der Waals volumes
	[H6v] H autocorrelation of lag 6 / weighted by atomic van der Waals volumes
	[H7v] H autocorrelation of lag 7 / weighted by atomic van der Waals volumes
	[H8v] H autocorrelation of lag 8 / weighted by atomic van der Waals volumes
	[R1v] R autocorrelation of lag 1 / weighted by atomic van der Waals volumes
	[R2v] R autocorrelation of lag 2 / weighted by atomic van der Waals volumes
	[R3v] R autocorrelation of lag 3 / weighted by atomic van der Waals volumes

	[R4v] R autocorrelation of lag 4 / weighted by atomic van der Waals volumes
	[R5v] R autocorrelation of lag 5 / weighted by atomic van der Waals volumes
	[R6v] R autocorrelation of lag 6 / weighted by atomic van der Waals volumes
	[R7v] R autocorrelation of lag 7 / weighted by atomic van der Waals volumes
	[R8v] R autocorrelation of lag 8 / weighted by atomic van der Waals volumes
	[R1v+] R maximal autocorrelation of lag 1 / weighted by atomic van der Waals volumes
	[R2v+] R maximal autocorrelation of lag 2 / weighted by atomic van der Waals volumes
	[R3v+] R maximal autocorrelation of lag 3 / weighted by atomic van der Waals volumes
	[R4v+] R maximal autocorrelation of lag 4 / weighted by atomic van der Waals volumes
	[R5v+] R maximal autocorrelation of lag 5 / weighted by atomic van der Waals volumes
	[R6v+] R maximal autocorrelation of lag 6 / weighted by atomic van der Waals volumes
	[R7v+] R maximal autocorrelation of lag 7 / weighted by atomic van der Waals volumes
	[R8v+] R maximal autocorrelation of lag 8 / weighted by atomic van der Waals volumes
	[HATSv] leverage-weighted total index / weighted by atomic van der Waals volumes
	[HTv] H total index / weighted by atomic van der Waals volumes
	[RTv] R total index / weighted by atomic van der Waals volumes
electronegativity weighted	
	[HATS0e] leverage-weighted autocorrelation of lag 0 / weighted by atomic Sanderson electronegativities
	[HATS1e] leverage-weighted autocorrelation of lag 1 / weighted by atomic Sanderson electronegativities
	[HATS2e] leverage-weighted autocorrelation of lag 2 / weighted by atomic Sanderson electronegativities
	[HATS3e] leverage-weighted autocorrelation of lag 3 / weighted by atomic Sanderson electronegativities
	[HATS4e] leverage-weighted autocorrelation of lag 4 / weighted by atomic Sanderson electronegativities
	[HATS5e] leverage-weighted autocorrelation of lag 5 / weighted by atomic Sanderson electronegativities
	[HATS6e] leverage-weighted autocorrelation of lag 6 / weighted by atomic Sanderson electronegativities
	[HATS7e] leverage-weighted autocorrelation of lag 7 / weighted by atomic Sanderson electronegativities
	[HATS8e] leverage-weighted autocorrelation of lag 8 / weighted by atomic Sanderson electronegativities
	[H0e] H autocorrelation of lag 0 / weighted by atomic Sanderson electronegativities
	[H1e] H autocorrelation of lag 1 / weighted by atomic Sanderson electronegativities
	[H2e] H autocorrelation of lag 2 / weighted by atomic Sanderson electronegativities

	[H3e] H autocorrelation of lag 3 / weighted by atomic Sanderson electronegativities
	[H4e] H autocorrelation of lag 4 / weighted by atomic Sanderson electronegativities
	[H5e] H autocorrelation of lag 5 / weighted by atomic Sanderson electronegativities
	[H6e] H autocorrelation of lag 6 / weighted by atomic Sanderson electronegativities
	[H7e] H autocorrelation of lag 7 / weighted by atomic Sanderson electronegativities
	[H8e] H autocorrelation of lag 8 / weighted by atomic Sanderson electronegativities
	[R1e] R autocorrelation of lag 1 / weighted by atomic Sanderson electronegativities
	[R2e] R autocorrelation of lag 2 / weighted by atomic Sanderson electronegativities
	[R3e] R autocorrelation of lag 3 / weighted by atomic Sanderson electronegativities
	[R4e] R autocorrelation of lag 4 / weighted by atomic Sanderson electronegativities
	[R5e] R autocorrelation of lag 5 / weighted by atomic Sanderson electronegativities
	[R6e] R autocorrelation of lag 6 / weighted by atomic Sanderson electronegativities
	[R7e] R autocorrelation of lag 7 / weighted by atomic Sanderson electronegativities
	[R8e] R autocorrelation of lag 8 / weighted by atomic Sanderson electronegativities
	[R1e+] R maximal autocorrelation of lag 1 / weighted by atomic Sanderson electronegativities
	[R2e+] R maximal autocorrelation of lag 2 / weighted by atomic Sanderson electronegativities
	[R3e+] R maximal autocorrelation of lag 3 / weighted by atomic Sanderson electronegativities
	[R4e+] R maximal autocorrelation of lag 4 / weighted by atomic Sanderson electronegativities
	[R5e+] R maximal autocorrelation of lag 5 / weighted by atomic Sanderson electronegativities
	[R6e+] R maximal autocorrelation of lag 6 / weighted by atomic Sanderson electronegativities
	[R7e+] R maximal autocorrelation of lag 7 / weighted by atomic Sanderson electronegativities
	[R8e+] R maximal autocorrelation of lag 8 / weighted by atomic Sanderson electronegativities
	[HATSe] leverage-weighted total index / weighted by atomic Sanderson electronegativities
	[HTe] H total index / weighted by atomic Sanderson electronegativities
	[RTe] R total index / weighted by atomic Sanderson electronegativities
polarizability weighted	
	[HATS0p] leverage-weighted autocorrelation of lag 0 / weighted by atomic polarizabilities
	[HATS1p] leverage-weighted autocorrelation of lag 1 / weighted by atomic polarizabilities
	[HATS2p] leverage-weighted autocorrelation of lag 2 / weighted by atomic polarizabilities

	[HATS3p] leverage-weighted autocorrelation of lag 3 / weighted by atomic polarizabilities
	[HATS4p] leverage-weighted autocorrelation of lag 4 / weighted by atomic polarizabilities
	[HATS5p] leverage-weighted autocorrelation of lag 5 / weighted by atomic polarizabilities
	[HATS6p] leverage-weighted autocorrelation of lag 6 / weighted by atomic polarizabilities
	[HATS7p] leverage-weighted autocorrelation of lag 7 / weighted by atomic polarizabilities
	[HATS8p] leverage-weighted autocorrelation of lag 8 / weighted by atomic polarizabilities
	[H0p] H autocorrelation of lag 0 / weighted by atomic polarizabilities
	[H1p] H autocorrelation of lag 1 / weighted by atomic polarizabilities
	[H2p] H autocorrelation of lag 2 / weighted by atomic polarizabilities
	[H3p] H autocorrelation of lag 3 / weighted by atomic polarizabilities
	[H4p] H autocorrelation of lag 4 / weighted by atomic polarizabilities
	[H5p] H autocorrelation of lag 5 / weighted by atomic polarizabilities
	[H6p] H autocorrelation of lag 6 / weighted by atomic polarizabilities
	[H7p] H autocorrelation of lag 7 / weighted by atomic polarizabilities
	[H8p] H autocorrelation of lag 8 / weighted by atomic polarizabilities
	[R1p] R autocorrelation of lag 1 / weighted by atomic polarizabilities
	[R2p] R autocorrelation of lag 2 / weighted by atomic polarizabilities
	[R3p] R autocorrelation of lag 3 / weighted by atomic polarizabilities
	[R4p] R autocorrelation of lag 4 / weighted by atomic polarizabilities
	[R5p] R autocorrelation of lag 5 / weighted by atomic polarizabilities
	[R6p] R autocorrelation of lag 6 / weighted by atomic polarizabilities
	[R7p] R autocorrelation of lag 7 / weighted by atomic polarizabilities
	[R8p] R autocorrelation of lag 8 / weighted by atomic polarizabilities
	[R1p+] R maximal autocorrelation of lag 1 / weighted by atomic polarizabilities
	[R2p+] R maximal autocorrelation of lag 2 / weighted by atomic polarizabilities
	[R3p+] R maximal autocorrelation of lag 3 / weighted by atomic polarizabilities
	[R4p+] R maximal autocorrelation of lag 4 / weighted by atomic polarizabilities
	[R5p+] R maximal autocorrelation of lag 5 / weighted by atomic polarizabilities
	[R6p+] R maximal autocorrelation of lag 6 / weighted by atomic polarizabilities
	[R7p+] R maximal autocorrelation of lag 7 / weighted by atomic polarizabilities
	[R8p+] R maximal autocorrelation of lag 8 / weighted by atomic polarizabilities
	[HATSp] leverage-weighted total index / weighted by atomic polarizabilities
	[HTp] H total index / weighted by atomic polarizabilities

	[RTp] R total index / weighted by atomic polarizabilities
Unweighted	
	[HATS0u] leverage-weighted autocorrelation of lag 0 / unweighted
	[HATS1u] leverage-weighted autocorrelation of lag 1 / unweighted
	[HATS2u] leverage-weighted autocorrelation of lag 2 / unweighted
	[HATS3u] leverage-weighted autocorrelation of lag 3 / unweighted
	[HATS4u] leverage-weighted autocorrelation of lag 4 / unweighted
	[HATS5u] leverage-weighted autocorrelation of lag 5 / unweighted
	[HATS6u] leverage-weighted autocorrelation of lag 6 / unweighted
	[HATS7u] leverage-weighted autocorrelation of lag 7 / unweighted
	[HATS8u] leverage-weighted autocorrelation of lag 8 / unweighted
	[H0u] H autocorrelation of lag 0 / unweighted
	[H1u] H autocorrelation of lag 1 / unweighted
	[H2u] H autocorrelation of lag 2 / unweighted
	[H3u] H autocorrelation of lag 3 / unweighted
	[H4u] H autocorrelation of lag 4 / unweighted
	[H5u] H autocorrelation of lag 5 / unweighted
	[H6u] H autocorrelation of lag 6 / unweighted
	[H7u] H autocorrelation of lag 7 / unweighted
	[H8u] H autocorrelation of lag 8 / unweighted
	[R1u] R autocorrelation of lag 1 / unweighted
	[R2u] R autocorrelation of lag 2 / unweighted
	[R3u] R autocorrelation of lag 3 / unweighted
	[R4u] R autocorrelation of lag 4 / unweighted
	[R5u] R autocorrelation of lag 5 / unweighted
	[R6u] R autocorrelation of lag 6 / unweighted
	[R7u] R autocorrelation of lag 7 / unweighted
	[R8u] R autocorrelation of lag 8 / unweighted
	[R1u+] R maximal autocorrelation of lag 1 / unweighted
	[R2u+] R maximal autocorrelation of lag 2 / unweighted
	[R3u+] R maximal autocorrelation of lag 3 / unweighted
	[R4u+] R maximal autocorrelation of lag 4 / unweighted
	[R5u+] R maximal autocorrelation of lag 5 / unweighted

	[R6u+] R maximal autocorrelation of lag 6 / unweighted
	[R7u+] R maximal autocorrelation of lag 7 / unweighted
	[R8u+] R maximal autocorrelation of lag 8 / unweighted
	[HATSu] leverage-weighted total index / unweighted
	[HTu] H total index / unweighted
	[RTu] R total index / unweighted

Surface area descriptors

The atoms in a molecule are viewed as spheres whose radii are the van der Waals radii. Therefore we can speak of the surface of a molecule as well. There are three types of surfaces which can be defined for a molecule.

van der Waals Surface Area of a molecule is the outer envelope (one that separates atoms from the exterior) of the union of van der Waal surfaces of the atoms in the molecule. Thus, the van der Waal surface area of a molecule is a union of spheres.

Solvent Accessible Surface Area of a molecule is the outer envelope of the union of solvent accessible surfaces of atoms in the molecule. The solvent accessible surface of an atom is the surface of a sphere whose radius is 1.4 Ang + the van der Waal radius of that atom. Thus, the solvent accessible surface area of a molecule is also a union of spheres.

Molecular Surface Area of a molecule is the surface of a solid that can be obtained as follows. Suppose that we place a molecule in 3D space and use a spherical erasor to wipeout the space wherever possible. Of course, we cannot take the erasor into the molecule. Thus, what remains after this operation is the portion of the space occupied by the van der Waal spheres that represent the atoms in the molecule, plus the space in the wedge between the atoms that are not reachable by the erasor. Thus, the Molecular Surface Area of a molecule is a union of spheres and tori.

The first two surface areas are being computed in sarchitect. Given below is how the van der Waal surface area for a molecule can be computed (the Solvent Accessible surface area computation is similar).

Consider the van der Waal Surface Area of a molecule. For each atom, we define the van der Waal sphere as the sphere with center at the center of the atom (given by x,y,z in .sdf inputs) and radius being the respective van der Waal radius. The van der Waal Surface of a molecule is the outer envelope of the van der Waal spheres of the atoms. Since the van der Waal spheres of atoms overlap, only a portion of van der Waal sphere of each atom enters into the van der Waal Surface Area of a molecule. Among the atoms that contribute to the van der Waal Surface Area of a molecule, some atoms can be hydrophobic, some others hydrophilic, some charged positive, some negative and some neither, some aromatic and so on. Thus, we can speak of partial and fractional surface areas of particular property as well. These constitute the Surface Area Descriptors.

We compute the Van der Waal and solvent accessible surface area of molecules by slicing the molecule along a chosen axis and finding the contribution of each of the atoms in that slice to the total surface area. The size of the

slice can be set by the user, with thinner slices being more accurate. Surface area for the individual atoms is calculated separately so that one can compute the total as well as the partial surface area with special properties like Hydrophobicity, polarity, etc.

The tool is tested for Oxygen molecule and the error is less than < 1 % when the slices are 0.3 Å thick.

[tSA] accessible surface area (ASA)
[tVSA] van der Waals surface area
[tPSA] polar accessible surface area
[tNPSA] nonpolar accessible surface area
[tPVSA] polar van der Waals surface area
[tNPVSA] nonpolar van der Waals surface area
[fPSA] fractional polar accessible surface area
[fNPSA] fractional nonpolar accessible surface area
[fPVSA] fractional polar van der Waals surface area
[fNPVSA] fractional nonpolar van der Waals surface area
[t+veSA] accessible surface area of atoms with positive charge
[t-veSA] accessible surface area of atoms with negative charge
[t+veVSA] van der Waals surface area of atoms with positive charge
[t-veVSA] van der Waals surface area of atoms with negative charge
[f+veSA] fractional accessible surface area of atoms with positive charge
[f-veSA] fractional accessible surface area of atoms with negative charge
[f+veVSA] fractional van der Waals surface area of atoms with positive charge
[f-veVSA] fractional van der Waals surface area of atoms with negative charge

Geometrical Descriptors

Aromaticity indices

Aromaticity indices are geometrical descriptors corresponding to the aromaticity of a molecule. The following aromaticity indices are computed in sarchitect.

RC Index- Jug RC Index (RCI)

This index is based on the ring current whose magnitude is determined by its weakest link in the ring which is defined as the bond with the minimum total bond order (Jug, 1983; Jug, 1984). That is,

$$RCI = \min_i \pi_{Li}$$

where, π_{Li} is the bond order of the bond L_i given by:

$$\pi_{Li} = (a / r_{xy}) - b$$

where,

a and *b* are constants (Refer [1] for the table of constants), and

r_{xy} is the geometric distance between atoms *i* and *j* that are the end points of the bond L_i.

HOMT

HOMT (HOMA total) is defined as the sum:

$$HOMT = \sum_{k} \alpha k . \sum_{b=1}^{B_{\pi k}} (r_k^{opt} - r_b)^2$$

where,

r_b is the actual bond length,

$B_{\pi k}$ is the number of aromatic bonds of type *k* (see the table below for the Bond-Types),

α_k is a constant for the kth type of aromatic bond (see the table below), and

r_k^{opt} is the typical aromatic bond length for the kth type of aromatic bond (see the table below).

Table

Harmonic Oscillator Model of Aromaticity index (HOMA)

index is based on the degree of alternation of single and double bonds, measuring the bond length deviations from optimal lengths attributed to the typical aromatic state (Kruszewski & Krygowski, 1972).

$$HOMA = 1 - HOMT / B_{\pi}$$

where,

$B\pi$ is the total number of aromatic bonds, and

HOMT is as defined above.

Balaban 3D index (J3D)

The 3D-Balaban Index is derived from the geometry distance matrix (hence, a 3D descriptor). The geometry matrix *G* is a square symmetric matrix where the ij^{th} entry is the Euclidean distance between the ith and the jth atoms. The geometric distance degree σ_i^G is the ith row sum in the geometry matrix *G* for each *i*, that is,

$$\sigma_i^G = \sum_{j=1}^{n} G_{ij}$$

Now we define *J3D* as follows:

$$J3D = (B/(C+1))\sum_{b \in B}(\sigma_i^G.\sigma_j^G)_b^{-1/2}),$$

where,

σ_i^G and σ_j^G are the geometric distance degrees of two adjascent atoms *i* and ***j*** connected by the bond *b*, and the sum runs over all the bonds *b* in the molecule, *B* is the total number of bonds in the molecule, and *C* is the cyclomatic number.

Radius of gyration (RGyr)

Radius of Gyration is a size descriptor for the distribution of atomic masses in a molecule (Tanford, 1961; Volkenstein, 1963), defined as:

$$R_{Gyr} = sqrt(sum_{i=1}^{n} m_i n_i^2 / MW)$$

where, r_i is the distance of the ith atom from the centre of mass of the molecule, m_i is the corresponding atomic mass, *n* is the atom number, and *MW*, the molecular weight.

The radius of gyration is a measure of the compactness of a molecule especially for long-chain molecules.

Potential Energy

Non Bonded

[Electro] electrostatic energy

The Electrostatic energy descriptor is given as:

$$E = q_1 q_2 / r$$

where, q_1, q_2 are the partial charges taken from the MMFF force field, and *r* is the distance between q_1 and q_2.

Sarchitect Table of Contents