Molecular Descriptors Guide

Description of the Molecular Descriptors Appearing in the Toxicity Estimation Software Tool

Version 1.0.2

Table of Contents

I. Constitutional Descriptors (45)

1. Molecular weight (MW)
2. Sum of atomic van der Waals volumes (scaled on Carbon atom) (Sv)
3. Sum of atomic Sanderson electronegativities (scaled on Carbon atom) (Se)
4. Sum of atomic polarizabilities (scaled on Carbon atom) (Sp)
5. Sum of Kier-Hall electrotopological states (Ss)
6. Average molecular weight (AMW)
7. Mean atomic van der Waals volume (scaled on Carbon atom) (Mv)
8. Mean atomic Sanderson electronegativity (scaled on Carbon atom) (Me)
9. Mean atomic polarizability (scaled on Carbon atom) (Mp)
10. Mean electrotopological state (Ms)
11. Number of atoms (nAT)
12. Number of non-H atoms (nSK)
13. Number of bonds (nBT)
14. Number of non-H bonds (nBO)
15. Number of multiple bonds (nBM)
16. Sum of conventional bond orders (H-depleted) (SCBO)
17. Aromatic ratio (ARR)
18. Number of rings (nCIC)
19. Number of circuits (nCIR)
20. Number of double bonds (nDB)
21. Number of triple bonds (nTB)
22. Number of aromatic bonds (nAB)
23. Number of Hydrogen atoms (nH)
24. Number of Carbon atoms (nC)
25. Number of Nitrogen atoms (nN)
26. Number of Oxygen atoms (nO)
27. Number of Phosphorous atoms (nP)
28. Number of Sulfur atoms (nS)
29. Number of Fluorine atoms (nF)
30. Number of Chlorine atoms (nCL)
31. Number of Bromine atoms (nBR)
32. Number of Iodine atoms (nI)
33. Number of Boron atoms (nB)
34. Number of halogen atoms (nX)
35. Number of 3-membered rings (nR03)
36. Number of 4-membered rings (nR04)
37. Number of 5-membered rings (nR05)
38. Number of 6-membered rings (nR06)
39. Number of 7-membered rings (nR07)
40. Number of 8-membered rings (nR08)
41. Number of 9-membered rings (nR09)
42. Number of 10-membered rings (nR10)
43. Number of 11-membered rings (nR11)
44. Number of 12-membered rings (nR12)
45. Number of benzene-like rings (nBnz)

The molecular weight (MW) is the sum of the molecular weights of the individual atoms:

$$MW = \sum_{i=1}^{A} MW_i \qquad (1)$$

Similarly the sum of the van der Waals volumes is given by

$$Sv = \sum_{i=1}^{A} V_i \qquad (2)$$

The sum of the electronegativities (Se), polarizabilities (Sp), and Kier-Hall electrotopological states (Ss) are given by a similar formula to equation 2.

The average molecular weight is given as follows:

$$AMW = MW / nAT \qquad (3)$$

where nAT is the number of atoms.

The mean volume (Mv) is calculated by dividing the sum of the van der Waals volumes by the number of atoms:

$$Mv = Sv / nAT \qquad (4)$$

The mean electronegativity (Me) and mean polarizability (Mp) are calculated using formulas similar to equation 4.

The mean electrotopological state (Ms) is calculated by dividing Ss by the number of non-hydrogen atoms (nSK):

$$Ms = Ss / nSK \qquad (5)$$

The number of bonds (nBT) is the total number of bonds (double and triple bonds are counted as two and three bonds, respectively).

The number of non-H bonds (nBO) is the number of bonds that do not involve hydrogen atoms.

The number of multiple bonds (nBM) is the number of bonds that have a bond order greater than one (note: aromatic bonds have a bond order of 1.5).

The sum of the conventional bond orders (SCBO) is the sum of the bond orders of bonds not involving hydrogens (again aromatic bonds have an order of 1.5).

The aromatic ratio (ARR) is the fraction of aromatic atoms in the hydrogen suppressed graph.

The number of rings (nCIC) is the number of simple rings in the molecules that aren't cycles. The number of rings can be determined by

$$nCIC = nBO - nSK + 1 \qquad (6)$$

The number of circuits (nCIR) includes both rings and circuits (a circuit is a larger loop around two or more rings). As an example naphthalene contains 3 circuits and 2 rings.

The number of double bonds (nDB) is equal to the number of non-aromatic double bonds.

The number of triple bonds (nTB) is simply the count of triple bonds in the molecule.

The number of aromatic bonds (nAB) is the number of bonds with a bond order of 1.5.

The number of hydrogen (nH), carbon (nC), nitrogen (nN), oxygen (nO), phosphorus (nP), sulfur (nS), fluorine (nF), chlorine (nCl), bromine (nBr), iodine (nI), and boron (nB) atoms are simply the total number of each of these types of atoms in the molecule.

The number of halogen atoms is the simply the sum of the counts of the halogen atoms:

$$nX = nF + nCL + nBR + nI \qquad (7)$$

The number of the rings of each size (*nR03, nR04… nR12*) is the count of all the *cycles* with the given number of atoms.

The number of benzene-like rings (nBnz) is the number of six-membered aromatic rings containing no heteroatoms.

II. Chi Connectivity Indices (46)

A. Simple

1. Simple zero order chi index (0χ or x0)
2. Simple 1st order chi index (1χ or x1)
3. Simple 2nd order chi index (2χ or x2)
4. Simple 3rd order path chi index ($3\chi p$ or xp3)
5. Simple 4th order path chi index ($4\chi p$ or xp4)
6. Simple 5th order path chi index ($5\chi p$ or xp5)
7. Simple 6th order path chi index ($6\chi p$ or xp6)
8. Simple 7th order path chi index ($7\chi p$ or xp7)
9. Simple 8th order path chi index ($8\chi p$ or xp8)
10. Simple 9th order path chi index ($9\chi p$ or xp9)
11. Simple 10th order path chi index ($10\chi p$ or xp10)
12. Simple 3rd order cluster chi index ($3\chi c$ or xc3)
13. Simple 4th order cluster chi index ($4\chi c$ or xc4)
14. Simple 4th order path/cluster chi index ($4\chi pc$ or xpc4)
15. Simple 3rd order chain chi index ($3\chi ch$ or xch3)
16. Simple 4th order chain chi index ($4\chi ch$ or xch4)
17. Simple 5th order chain chi index ($5\chi ch$ or xch5)
18. Simple 6th order chain chi index ($6\chi ch$ or xch6)
19. Simple 7th order chain chi index ($7\chi ch$ or xch7)
20. Simple 8th order chain chi index ($8\chi ch$ or xch8)
21. Simple 9th order chain chi index ($9\chi ch$ or xch9)
22. Simple 10th order chain chi index ($10\chi ch$ or xch10)
23. Difference between chi cluster-3 and chi path/cluster-4 (knotp)

The general formula for the simple chi connectivity indices ($^m\chi_t$ or xtm) is as follows:

$$^m\chi_t = \sum_{i=1}^{A} {}^m c_i \tag{1}$$

where m = bond order (0 = atoms, 1 = fragments of one bond, 2 = fragments of two bonds, etc.), t = type of calculation (p = path, c = cluster, pc = path/cluster, ch = chain or cycle) and A = the number of non-hydrogen atoms in the molecule. For the first three path indices (x0, x1, and x2), the calculation type, p, is omitted from the variable names used in the software.

$^m c_i$ is calculated for all of the fragments of type t and path length m in the hydrogen depleted graph of the molecule:

$$^m c_i = \prod_{k=1}^{m+1} (\delta_k)^{-0.5} \tag{2}$$

where k = the different atoms in the fragment and δ_k is the vertex degree of an atom given by

$$\delta_k = \sigma_k - h_k \tag{3}$$

where σ_k is the number of electrons in sigma orbitals and h_k is the number of bonded hydrogen atoms (Simply $-\delta_k$ is the # of non-hydrogen atoms bonded to atom k).

knotp is simply the difference between xc3 and xpc4:
$$knotp = xc3 - xpc4 \qquad (4)$$
The simple chi indices are described on page 85 of the Handbook of Molecular Descriptors (Todeschini and Consonni 2000).

B. Valence

1. Valence zero order chi index ($0\chi v$ or xv0)
2. Valence 1st order chi index ($1\chi v$ or xv1)
3. Valence 2nd order chi index ($2\chi v$ or xv2)
4. Valence 3rd order path chi index ($3\chi vp$ or xvp3)
5. Valence 4th order path chi index ($4\chi vp$ or xvp4)
6. Valence 5th order path chi index ($5\chi vp$ or xvp5)
7. Valence 6th order path chi index ($6\chi vp$ or xvp6)
8. Valence 7th order path chi index ($7\chi vp$ or xvp7)
9. Valence 8th order path chi index ($8\chi vp$ or xvp8)
10. Valence 9th order path chi index ($9\chi vp$ or xvp9)
11. Valence 10th order path chi index ($10\chi vp$ or xvp10)
12. Valence 3rd order cluster chi index ($3\chi vc$ or xvc3)
13. Valence 4th order cluster chi index ($4\chi vc$ or xvc4)
14. Valence 4th order path/cluster chi index ($4\chi vpc$ or xvpc4)
15. Valence 3rd order chain chi index ($3\chi vch$ or xvch3)
16. Valence 4th order chain chi index ($4\chi vch$ or xvch4)
17. Valence 5th order chain chi index ($5\chi vch$ or xvch5)
18. Valence 6th order chain chi index ($6\chi vch$ or xvch6)
19. Valence 7th order chain chi index ($7\chi vch$ or xvch7)
20. Valence 8th order chain chi index ($8\chi vch$ or xvch8)
21. Valence 9th order chain chi index ($9\chi vch$ or xvch9)
22. Valence 10th order chain chi index ($10\chi vch$ or xvch10)
23. Difference between chi valence cluster-3 and chi valence path/cluster-4 (knotpv)

The valence connectivity indices ($^m\chi^v_t$ or xvtm) are calculated in the same fashion as the simple connectivity indices except that the vertex degree in equation B2 is replaced by the valence vertex degree (δ^v) to give
$$^m c_i = \prod_{k=1}^{m+1} \left(\delta_k^v\right)^{-0.5} \qquad (5)$$
where the valence vertex degree is given by
$$\delta_k^v = Z_k^v - h_k = \sigma_k + \pi_k + n_k - h_k \qquad (6)$$
where Z_k^v is the number of valence electrons, π_k = number of electrons in pi orbitals, and n_k is the number of electrons in lone-pair orbitals.

For atoms of higher principal quantum levels, the valence vertex degree is given by

$$\delta_k^v = (Z_k^v - h_k)/(Z_k - Z_k^v - 1) \tag{7}$$

where Z_k is the number of electrons in atom k (the atomic number).

The valence chi indices are described on page 86 of the Handbook of Molecular Descriptors (Todeschini and Consonni 2000). The chi connectivity indices are described in further detail in the literature (Elsevier_MDL 2006; Hall and Kier 1991, 1999; Kier and Hall 1976, 1986).

III. Kappa Shape Indices (7)

1. First order kappa shape index ($^1\kappa$ or k1)
2. Second order kappa shape index ($^2\kappa$ or k2)
3. Third order kappa shape index ($^3\kappa$ or k3)
4. First order kappa alpha shape index ($^1\kappa\alpha$ or ka1)
5. Second order kappa alpha shape index ($^2\kappa\alpha$ or ka2)
6. Third order kappa alpha shape index ($^3\kappa\alpha$ or ka3)
7. Kappa flexibility index (phia)

The first order kappa shape index ($^1\kappa$ or k1) is given by

$$^1\kappa = \frac{A(A-1)^2}{\left(^1P\right)^2} \tag{1}$$

where ^iP = # of paths of bond length i in the hydrogen suppressed molecule and A is the number of non hydrogen atoms in the molecule.

The second order kappa shape index ($^2\kappa$ or k2) is given by

$$^2\kappa = \frac{(A-1)(A-2)^2}{\left(^2P\right)^2} \tag{2}$$

The third order kappa shape index ($^3\kappa$ or k3) is given by

$$^3\kappa = \frac{(A-1)(A-3)^2}{\left(^3P\right)^2} \quad \text{if A is odd} \tag{3}$$

$$^3\kappa = \frac{(A-3)(A-2)^2}{\left(^3P\right)^2} \quad \text{if A is even} \tag{4}$$

The kappa shape indices are described on page 248 of the Handbook of Molecular descriptors (Todeschini and Consonni 2000).

The first order kappa alpha shape index ($^1\kappa\alpha$ or ka1) is given by

$$^1\kappa_\alpha = \frac{(A+\alpha)(A+\alpha-1)^2}{\left(^1P+\alpha\right)^2} \tag{5}$$

where

$$\alpha = 1 - \frac{r_x}{r_{x(sp^3)}} \tag{6}$$

where r_x is the covalent radius of the atom being evaluated and $r_{x(sp_3)}$ is the covalent radius of a carbon sp^3 atom (0.77 Å).

The covalent radii used in the software are as follows:

Atom / Hybrid	R (Å)	Atom / Hybrid	R (Å)
Csp^3	0.77	F	0.72
Csp^2	0.67	Cl	0.99
Csp	0.60	Br	1.14
Nsp^3	0.74	I	1.33
Nsp^2	0.62		
Nsp	0.55	Si	1.17
Osp^3	0.74	As	1.22
Osp^2	0.62	Sn	1.42
Psp^3	1.10		
Psp^2	1.00	Hg	1.48
Ssp^3	1.04	Pb	1.54
Ssp^2	0.94		

The second order kappa alpha shape index ($^2\kappa\alpha$ or ka2) is given by

$$^2\kappa_\alpha = \frac{(A+\alpha-1)(A+\alpha-2)^2}{(^2P+\alpha)^2}$$

(7)

The third order kappa alpha shape index ($^3\kappa\alpha$ or ka3) is given by

$$^3\kappa_\alpha = \frac{(A+\alpha-1)(A+\alpha-3)^2}{(^3P+\alpha)^2} \text{ if A is odd}$$

(8)

$$^3\kappa_\alpha = \frac{(A+\alpha-3)(A+\alpha-2)^2}{(^3P+\alpha)^2} \qquad \text{if A is even}$$

(9)

The kappa shape indices are described on page 250 of the Handbook of Molecular Descriptors (Todeschini and Consonni 2000).

The kappa flexibility index (phia) is given by

$$phia = \frac{^1\kappa_\alpha \, ^2\kappa_\alpha}{A}$$

(10)

The kappa flexibility index is described on page 178 of the Handbook of Molecular Descriptors (Todeschini and Consonni 2000).

IV. Electrotopological State Indices (166)

A. Atom Type E-State Sum Non-hydrogen Indices

1. Sum of (– CH3) E-States (SsCH3)
2. Sum of (= CH2) E-States (SdCH2)
3. Sum of (– CH2 –) E-States (SssCH2)
4. Sum of (tCH) E-States (StCH)
5. Sum of (= CH –) E-States (SdsCH)
6. Sum of (aaCH) E-States (SaaCH)
7. Sum of (> CH –) E-States (SsssCH)
8. Sum of (= C =) E-States (SddC)
9. Sum of (tC –) E-States (StsC)
10. Sum of (= C <) E-States (SdssC)
11. Sum of (saaC) E-States (SsaaC)
12. Sum of (aaaC) E-States (SaaaC)
13. Sum of (> C <) E-States (SssssC)
14. Sum of (– NH2) E-States (SsNH2)
15. Sum of (= NH) E-States (SdNH)
16. Sum of (– NH –) E-States (SssNH)
17. Sum of (aaNH) E-States (SaaNH)
18. Sum of (tN) E-States (StN)
19. Sum of (= N –) E-States (SdsN)
20. Sum of (aaN) E-States (SaaN)
21. Sum of (> N –) E-States (SsssN)
22. Sum of (aaaN) E-States (SaaaN)
23. Sum of (saaN) E-States (SsaaN)
24. Sum of (– NH3+) E-States (SsNH3p)
25. Sum of (– NH2+ –) E-States (SssNH2p)
26. Sum of (> NH+ –) E-States (SsssNHp)
27. Sum of (= N+ =) E-States (SddNp)
28. Sum of (# N+ –) E-States (StsNp)
29. Sum of (= N -) E-States (SdNm)
30. Sum of (= N+ <) E-States (SdssNp)
31. Sum of (> N+ <) E-States (SssssNp)
32. Sum of (saaN+) E-States (SsaaNp)
33. Sum of (– OH) E-States (SsOH)
34. Sum of (= O) E-States (SdO)
35. Sum of (– O –) E-States (SssO)
36. Sum of (aaO) E-States (SaaO)
37. Sum of (– O-) E-States (SsOm)
38. Sum of (– F) E-States (SsF)
39. Sum of (– Cl) E-States (SsCl)
40. Sum of (– Br) E-States (SsBr)
41. Sum of (– I) E-States (SsI)
42. Sum of (– SiH3) E-States (SsSiH3)
43. Sum of (– SiH2 –) E-States (SssSiH2)
44. Sum of (> SiH –) E-States (SsssSiH)
45. Sum of (> Si <) E-States (SssssSi)
46. Sum of (> P –) E-States (SsssP)
47. Sum of (dsssP) E-States (SdsssP)
48. Sum of (ddsP) E-States (SddsP)
49. Sum of (sssssP) E-States (SsssssP)
50. Sum of (=PH<) E-States (SdssPH)
51. Sum of (– PH2) E-States (SsPH2)
52. Sum of (– PH –) E-States (SssPH)
53. Sum of (– SH) E-States (SsSH)
54. Sum of (= S) E-States (SdS)
55. Sum of (– S –) E-States (SssS)
56. Sum of (aaS) E-States (SaaS)
57. Sum of (= S <) E-States (SdssS)
58. Sum of (ddssS) E-States (SddssS)
59. Sum of (sssssssS) E-States (SsssssssS)
60. Sum of (– SnH3) E-States (SsSnH3)
61. Sum of (– SnH2 –) E-States (SssSnH2)
62. Sum of (> SnH –) E-States (SsssSnH)
63. Sum of (> Sn <) E-States (SssssSn)
64. Sum of (– PbH3) E-States (SsPbH3)
65. Sum of (– PbH2 –) E-States (SssPbH2)
66. Sum of (> PbH –) E-States (SsssPbH)
67. Sum of (> Pb <) E-States (SssssPb)
68. Sum of (– As <) E-States (SsssAs)
69. Sum of (dsssAs) E-States (SdsssAs)
70. Sum of (ddsAs) E-States (SddsAs)
71. Sum of (sssssAs) E-States (SsssssAs)
72. Sum of (– AsH2) E-States (SsAsH2)
73. Sum of (– AsH –) E-States (SssAsH)
74. Sum of (– Hg+) E-States (SsHgp)
75. Sum of (– Hg –) E-States (SssHg)
76. Sum of (– Hg) E-States (SsHg)

The E-State value for a given non hydrogen atom i in a molecule is given by its intrinsic state (I_i) plus the sum of the perturbations on that atom by all the other atoms in the molecule:

$$S_k = I_k + \sum_{i=1}^{A} \Delta I_{ki} \qquad (1)$$

where the intrinsic state (I_k) is given by

$$I_k = \frac{(2/N)^2 \delta_k^v + 1}{\delta_k} \qquad (2)$$

where N = principal quantum number (which is equal to the element's period or row in the periodic table) where δ_k^v is calculated using equation 6 from Section II (even for elements with higher principal quantum levels).

The perturbation of atom k due to atom i is given by

$$\Delta I_{ki} = \frac{(I_i - I_k)}{r_{ki}^2} \qquad (4)$$

where

$$r_{ki} = d_{ki} + 1 \qquad (5)$$

where d_{ki} = is the number of bonds that separate atom k from atom i.

The atom type non-hydrogen indices (SX) are obtained by summing the E-State values for all the atoms of a given type t that are present in the molecule:

$$SX = \sum S(t) \qquad (6)$$

where S(t) is the atom E-State value (obtained from equation 1) for an atom of type t.

For example, for propane, the E-States for the individual atoms are as follows:

S(2)=1.25

H$_3$C   CH$_3$

S(1)=2.125   S(3)=2.125
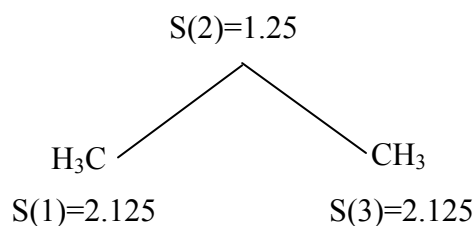
SsCH3 (the sum of the E-States for carbon atoms with three hydrogens attached and one single bond to a nonhydrogen atom) is 4.25 (2× 2.125). SssCH2 (the sum of the E-States for carbon atoms with two hydrogens attached and two single bonds to a nonhydrogen atoms) is 1.25.

B. Atom Type Count Non-hydrogen Indices

1. Count of (– CH3) (SsCH3_acnt)
2. Count of (= CH2) (SdCH2_acnt)
3. Count of (– CH2 –) (SssCH2_acnt)
4. Count of (tCH) (StCH_acnt)
5. Count of (= CH –) (SdsCH_acnt)
6. Count of (aaCH) (SaaCH_acnt)
7. Count of (> CH –) (SsssCH_acnt)
8. Count of (= C =) (SddC_acnt)
9. Count of (tC –) (StsC_acnt)
10. Count of (= C <) (SdssC_acnt)
11. Count of (saaC) (SsaaC_acnt)
12. Count of (aaaC) (SaaaC_acnt)
13. Count of (> C <) (SssssC_acnt)
14. Count of (– NH2) (SsNH2_acnt)
15. Count of (= NH) (SdNH_acnt)
16. Count of (– NH –) (SssNH_acnt)
17. Count of (aaNH) (SaaNH_acnt)
18. Count of (tN) (StN_acnt)
19. Count of (= N –) (SdsN_acnt)
20. Count of (aaN) (SaaN_acnt)
21. Count of (> N –) (SsssN_acnt)
22. Count of (aaaN) (SaaaN_acnt)
23. Count of (saaN) (SsaaN_acnt)
24. Count of (– NH3+) (SsNH3p_acnt)
25. Count of (– NH2+ –) (SssNH2p_acnt)
26. Count of (> NH+ –) (SsssNHp_acnt)
27. Count of (= N+ =) (SddNp_acnt)
28. Count of (# N+ –) (StsNp_acnt)
29. Count of (= N -) (SdNm_acnt)
30. Count of (= N+ <) (SdssNp_acnt)
31. Count of (> N+ <) (SsssNp_acnt)
32. Count of (saaN+) (SsaaNp_acnt)
33. Count of (– OH) (SsOH_acnt)
34. Count of (= O) (SdO_acnt)
35. Count of (– O –) (SssO_acnt)
36. Count of (aaO) (SaaO_acnt)
37. Count of (– O-) (SsOm_acnt)
38. Count of (– F) (SsF_acnt)
39. Count of (– Cl) (SsCl_acnt)
40. Count of (– Br) (SsBr_acnt)
41. Count of (– I) (SsI_acnt)
42. Count of (– SiH3) (SsSiH3_acnt)
43. Count of (– SiH2 –) (SssSiH2_acnt)
44. Count of (> SiH –) (SsssSiH_acnt)
45. Count of (> Si <) (SssssSi_acnt)
46. Count of (> P –) (SsssP_acnt)
47. Count of (dsssP) (SdsssP_acnt)
48. Count of (ddsP) (SddsP_acnt)
49. Count of (sssssP) (SssssssP_acnt)
50. Count of (=PH<) (SdssPH_acnt)
51. Count of (– PH2) (SsPH2_acnt)
52. Count of (– PH –) (SssPH_acnt)
53. Count of (– SH) (SsSH_acnt)
54. Count of (= S) (SdS_acnt)
55. Count of (– S –) (SssS_acnt)
56. Count of (aaS) (SaaS_acnt)
57. Count of (= S <) (SdssS_acnt)
58. Count of (ddssS) (SddssS_acnt)
59. Count of (sssssssS) (SsssssssS_acnt)
60. Count of (– SnH3) (SsSnH3_acnt)
61. Count of (– SnH2 –) (SssSnH2_acnt)
62. Count of (> SnH –) (SsssSnH_acnt)
63. Count of (> Sn <) (SssssSn_acnt)
64. Count of (– PbH3) (SsPbH3_acnt)
65. Count of (– PbH2 –) (SssPbH2_acnt)
66. Count of (> PbH –) (SsssPbH_acnt)
67. Count of (> Pb <) (SssssPb_acnt)
68. Count of (– As <) (SsssAs_acnt)
69. Count of (dsssAs) (SdsssAs_acnt)
70. Count of (ddsAs) (SddsAs_acnt)
71. Count of (sssssAs) (SsssssAs_acnt)
72. Count of (– AsH2) (SsAsH2_acnt)
73. Count of (– AsH –) (SssAsH_acnt)
74. Count of (– Hg+) (SsHgp_acnt)
75. Count of (– Hg –) (SssHg_acnt)
76. Count of (– Hg) (SsHg_acnt)

The atom type count non-hydrogen indices are simply the count of each type of e-state fragment within a molecule. For the propane example, SsCH3_acnt (the count of all – CH3 groups in molecule) is 2 and SssCH2_acnt (the count of all – CH2 – groups in a molecule) is 1.

C. Hydrogen Indices

1. Sum of ([− OH]) hydrogen E-States (SHsOH)
2. Sum of ([= NH]) hydrogen E-States (SHdNH)
3. Sum of ([− SH]) hydrogen E-States (SHsSH)
4. Sum of ([− NH2]) hydrogen E-States (SHsNH2)
5. Sum of ([− NH −]) hydrogen E-States (SHssNH)
6. Sum of ([tCH]) hydrogen E-States (SHtCH)
7. Sum of (CH or CH2 with -F or -Cl attached) hydrogen E-State values (SHCHnX)
8. Sum of hydrogen E-State values for other atom types (SHother)

The hydrogen electrotopological state index (HS_i) for atom number i is given by

$$HS_i = KHE_i + \left(KHE_i - KHE(H_i)\right) + \sum_{j \neq i} \frac{KHE_j - KHE(H_i)}{\left(d_{ij} + 1\right)^2} \tag{7}$$

where KHE_i is the Kier-Hall electronegativity of the ith heavy atom in the H-depleted graph, $KHE(H_i)$ is the Kier-Hall electronegativity of the attached hydrogen atom (taken to be -0.2), and d_{ij} is the distance between each j heavy atom and the ith heavy atom to which the hydrogen is bonded. The Kier-Hall electronegativity of a given atom is given by

$$KHE_i = \frac{\delta_i^v - \delta_i}{N_i^2} \tag{8}$$

where in this case δ_i^v is the valence vertex degree given in equation 6 of Section II and N_i is the principal quantum number of atom i.

SHsOH is the sum of the hydrogen E-States for the atoms with an E-State type of SsOH.

SHdNH is the sum of the hydrogen E-States for the atoms with an E-State type of SdNH.

SHsSH is the sum of the hydrogen E-States for the atoms with an E-State type of SsSH.

SHsNH2 is the sum of the hydrogen E-States for the atoms with an E-State type of SsNH2.

SHssNH is the sum of the hydrogen E-States for the atoms with an E-State type of SssNH or SaaNH.

SHtCH is the sum of the hydrogen E-States for the atoms with an E-State type of StCH.

SHCHnX is the sum of the hydrogen E-States for the atoms with an E-State type of SsCH3, SssCH2, SsssCH, or SdsCH when there is a chlorine or fluorine atom attached.

SHother is the sum of the hydrogen E-States for atoms (that have hydrogens attached) that do not fall into the other hydrogen E-State types listed above.

D. Maximum and Minimum Values

1. Maximum hydrogen E-State value in molecule (Hmax)
2. Maximum E-State value in molecule (Gmax)
3. Minimum hydrogen E-State value in molecule (Hmin)
4. Minimum E-State value in molecule (Gmin)
5. Maximum positive hydrogen E-State value in molecule. (Hmaxpos)
6. Minimum negative hydrogen E-State value in molecule. (Hminneg)

Hmax is the maximum hydrogen E-state value (calculated using equation 7) for all the atoms in the molecule.

Gmax is the maximum E-state value (calculated using equation 1) for all the atoms in the molecule.

Hmin is the minimum hydrogen E-state value for all the atoms in the molecule.

Gmin is the minimum E-state value for all the atoms in the molecule.

Hmaxpos is the maximum positive hydrogen E-state value for all the atoms in the molecule.

Hminneg is the minimum negative hydrogen E-state value for all the atoms in the molecule.

The calculation of E-states indices is described in further detail by Kier and Hall(Kier and Hall 1999).

V. Topological descriptors (15)

1.   First Zagreb index weighted by vertex degrees (ZM1)
2.   First Zagreb index weighted by valence vertex degrees (ZM1V)
3.   Second Zagreb index weighted by vertex degrees (ZM2)
4.   Second Zagreb index weighted by valence vertex degrees (ZM2V)
5.   Balaban distance connectivity index (J)
6. J_t index (J_t)
7.   Balaban centric index (BAC)
8.   Lopping centric index (Lop)
9.   Radial centric information index (ICR)
10.   E-State topological parameter (TIE)
11.   Maximal electrotopological negative variation (MAXDN)
12.   Maximal electrotopological positive variation (MAXDP)
13.   Molecular electrotopological variation (DELS)
14.   Wiener Index (W)
15.   Mean Wiener Index (WA)

The first Zagreb index (weighted by vertex degrees) is given by

$$ZM1 = \sum_a \delta_a^{\;2} \tag{1}$$

where a runs over the A atoms of the molecule and δ_a is the vertex degree of atom a.

$$ZM1V = \sum_a \delta_a^{v2} \tag{2}$$

where δ_a^v is the valence vertex degree of atom a.

$$ZM2 = \sum_b \left(\delta_i \cdot \delta_j \right)_b \tag{3}$$

where b runs over all of the bonds in the molecule.

$$ZM2V = \sum_b \left(\delta_i^v \cdot \delta_j^v \right)_b \tag{4}$$

The Zagreb indices (ZM1, ZM1V, ZM2, and ZM2V) are described on pg 509 of the Handbook of Molecular descriptors (Todeschini and Consonni 2000).

The Balaban distance connectivity index (J) is given by

$$J = \frac{B}{C+1} \sum_b \left(\sigma_i \sigma_j \right)_b^{-1/2} \tag{5}$$

where σ_i and σ_j are the vertex distance degrees of two adjacent atoms, and the sum runs over all the molecular bonds b; B is the number of bonds, and C (the cyclomatic number) is the number of rings. The vertex distance degree (σ_i) is the sum of the distances of all the other atoms to a given atom:

$$\sigma_i = \sum_{j=1}^{A} d_{ij} \tag{6}$$

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

The cyclomatic number is given by:
$$C = B - A + 1 \qquad (7)$$

A similar index, the Jt index, is defined as follows:
$$J_t = \frac{B}{C+1} \sum_b \left(t_i t_j \right)_b^{-1/2} \qquad (8)$$

where t_i is a different local vertex invariant given by
$$t_i = \frac{\sigma_i}{\delta_i} \qquad (9)$$

J and Jt are described on pg 21 of the Handbook of Molecular descriptors (Todeschini and Consonni 2000).

To calculate the centric indices, the vertices with a vertex degree of unity are stepwise removed until no more vertices can be removed. The Balaban centric index is calculated as follows:
$$BAC = \sum_{g=1}^{R} n_g^{\;2} \qquad (10)$$

where n_g is the number of atoms with a vertex degree of unity removed at step g and R is the number of removal steps.

The Lopping centric index is calculated as follows:
$$Lop = -\sum_{i=1}^{R} \frac{n_g}{A} \log\left(\frac{n_g}{A} \right) \qquad (11)$$

BAC and Lop are described on pg 42 of the Handbook of Molecular descriptors(Todeschini and Consonni 2000).

The radial centric information index (ICR) is calculated as follows:
$$ICR = -\sum_{g-1}^{G} \frac{n_g}{A} \log_2\left(\frac{n_g}{A} \right) \qquad (12)$$

where n_g is the number of vertices having equal atom eccentricities (η) in the gth class, G is the number of equivalence classes, and A is the atom number. The atom eccentricity is the maximum distance from the ith vertex to the other vertices:
$$\eta_i = \max_j \left(d_{ij} \right) \qquad (13)$$

ICR is described on pg 44 of the Handbook of Molecular descriptors (Todeschini and Consonni 2000).

Voeckel (1994) defines the E-State topological parameter (TIE) as follows:
$$TIE = \frac{B}{C+1} \sum_{b=1}^{B} \left(S_i \cdot S_j \right)_b^{-1/2} \qquad (14)$$

In order to match the descriptor used in Dragon (Talete 2006) (and to be able to calculate TIE for compounds with atoms with negative E-State values) the E-State topological parameter is instead given by:

$$TIE = \frac{B}{C+1} \sum_{b=1}^{B} \left(1 + \exp(S_i) \cdot \exp(S_j)\right)_b^{-1/2} \tag{15}$$

where B is the number of bonds, C is the cyclomatic number, S_i is the E-State for atom i. TIE is described on pg 165 of the Handbook of Molecular descriptors (Todeschini and Consonni 2000).

The maximal electrotopological negative variation (MAXDN) is given by

$$MAXDN = \max_i |\Delta I_i| \quad if \ \Delta I_i < 0 \quad i = 1, \ldots A \tag{16}$$

where ΔI_i is the field effect on the ith atom due to the perturbation of all other atoms:

$$\Delta I_i = \sum_j \frac{I_i - I_j}{\left(d_{ij} + 1\right)^2} \tag{17}$$

Similarly the maximal electrotopological positive variation (MAXDP) is given by

$$MAXDP = \max_i |\Delta I_i| \quad if \ \Delta I_i > 0 \quad i = 1, \ldots A \tag{18}$$

The molecular electrotopological variation (*DELS*) is given by

$$DELS = \sum_i |\Delta I_i| \tag{19}$$

MAXDN, MAXDP, and DELS are described in more detail by Gramatica (Gramatica et al. 2000).

The Wiener index (*W*) is given by

$$W = \frac{1}{2} \sum_{i=1}^{A} \sum_{i=1}^{A} d_{ij} \tag{20}$$

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

The mean Wiener index is given by

$$WA = \frac{2W}{A(A-1)} \tag{21}$$

W and WA are described in more detail on pg 497 of the Handbook of Molecular descriptors (Todeschini and Consonni 2000).

VI. Molecular Fragments (234)

1. ACAC
2. AN [attached to AN in same ring]
3. ANH [attached to AN in same ring]
4. C=C=O
5. C=[N+][O-]
6. C=[N+]=(Aguiara et al.)
7. -N=S=O
8. -N=[N+]=(Aguiara et al.)
9. -[N+]#N
10. A[N+][O-]
11. N+ [four single bonds]
12. CH2=C(CH3)C(=O)O-
13. CH2=CHC(=O)O-
14. -N=C=S
15. -SC(=S)S-
16. -SC(=S)O-
17. -SC(=O)S-
18. -SC(=O)O-
19. -OC(=S)O-
20. -C(=O)OC(=O)-
21. -NH- [3 membered ring]
22. -N< [3 membered ring]
23. -NH2 [attached to P]
24. -NH- [attached to P]
25. -N< [attached to P]
26. Sn=O
27. Sn [oxygen attach]
28. Sn [aromatic attach]
29. Sn
30. Si [aromatic attach]
31. Si [oxygen, aromatic attach]
32. Si [oxygen attach]
33. Si
34. -S(=O)(=O)- [2 nitrogen attach]
35. -S(=O)(=O)- [nitrogen, aromatic attach]
36. -S(=O)(=O)- [aromatic attach]
37. -S(=O)(=O)- [olefinic attach]
38. -S(=O)(=O)- [nitrogen, aliphatic attach]
39. -S(=O)(=O)- [aliphatic attach]
40. -S(=O)- [2 nitrogen attach]
41. -S(=O)- [nitrogen, aromatic attach]
42. -S(=O)- [aromatic attach]
43. -S(=O)- olefinic attach]
44. -S(=O)- [nitrogen, aliphatic attach]
45. -S(=O)- [aliphatic attach]
46. -OH [aromatic attach]
47. -SH [aromatic attach]
48. -N=O [aromatic attach]
49. -ONO2
50. -NO2 [aromatic attach]
51. -NO2 [olefinic attach]
52. -NO2 [nitrogen attach]
53. -NO2 [aliphatic attach]
54. -N=C=O [aromatic attach]
55. -N=C=O [aliphatic attach]
56. -COOH [aromatic attach]
57. -OC(=O)O-
58. -C(=O)O- [aromatic attach]
59. -C(=O)SH [aromatic attach]
60. -C(=O)S- [aromatic attach]
61. -C(=S)SH [aromatic attach]
62. -C(=S)S- [aromatic attach]
63. -C(=S)OH [aromatic attach]
64. -C(=S)O- [aromatic attach]
65. -C(=S)- [nitrogen, aromatic attach]
66. -C(=S)- [aromatic attach]
67. -CHO [aromatic attach]
68. -COOH [aliphatic attach]
69. HC(=O)O-
70. -C(=O)O- [nitrogen attach]
71. -C(=O)O- [cyclic]
72. -C(=O)O- [olefinic attach]
73. -C(=O)O- [aliphatic attach]
74. -C(=O)SH [nitrogen attach]
75. -C(=O)SH [aliphatic attach]
76. -C(=O)S- [nitrogen attach]
77. -C(=O)S- [aliphatic attach]
78. -C(=S)SH [nitrogen attach]
79. -C(=S)SH [aliphatic attach]
80. -C(=S)OH [aliphatic attach]
81. -C(=S)O- [nitrogen attach]
82. -C(=S)O- [aliphatic attach]
83. -C(=S)S- [nitrogen attach]
84. -C(=S)S- [aliphatic attach]
85. -C(=S)- [2 nitrogen attach]
86. -C(=S)- [nitrogen, aliphatic attach]

87. -C(=S)- [aliphatic attach]
88. -CF3 [aromatic attach]
89. -CCl3 [aromatic attach]
90. -C#N [aromatic attach]
91. -CH=NOH
92. -CH=NO
93. >C=NOH
94. >C=NO
95. -C([H])=N[H] [Nitrogen attach]
96. -C([H])=N [Nitrogen attach]
97. >C=N[H] [2 Nitrogen attach]
98. >C=N[H] [Nitrogen attach]
99. >C(=N) [2 Nitrogen attach]
100. >C(=N) [Nitrogen attach]
101. -CH=NH
102. >C=NH
103. CH2=N
104. -CH=N
105. >C=N
106. -NHN=O
107. >NN=O
108. -N(=O)
109. -N=NH
110. -N=N-
111. -NH2 [nitrogen attach]
112. -NH- [nitrogen attach]
113. -N< [nitrogen attach]
114. -CHO [nitrogen attach]
115. -CHO [aliphatic attach]
116. -C(=O)- [halogen attach]
117. -C(=O)- [2 aromatic attach]
118. -C(=O)- [nitrogen, aromatic attach]
119. -C(=O)- [aromatic attach]
120. -C(=O)- [2 nitrogen attach]
121. -C(=O)- [nitrogen, aliphatic attach]
122. -C(=O)- [phosphorus attach]
123. -C(=O)- [olefinic attach]
124. C=O(ketone, aliphatic attach)
125. C=O(non-ketone, aliphatic attach)
126. -CCl3 [aliphatic attach]
127. -CF3 [aliphatic attach]
128. -C#N [aliphatic nitrogen attach]
129. -C#N [aliphatic sulfur attach]
130. -C#N [olefinic attach]
131. -C#N [aliphatic attach]
132. P=NH
133. P=N
134. P=S
135. P(=O)
136. As(=O)
137. -SH [aliphatic attach]
138. -S- [three membered ring]
139. -S- [sulfur attach]
140. -S- [arsenic attach]
141. -S- [phosphorus attach]
142. -S- [nitrogen attach]
143. -S- [2 aromatic attach]
144. -S- [aromatic attach]
145. -OH [sulfur attach]
146. -OH [oxygen attach]
147. -OH [arsenic attach]
148. -OH [phosphorus attach]
149. -OH [nitrogen attach]
150. -OH [aliphatic attach]
151. -O- (epoxide)
152. -O- [2 aromatic attach]
153. -O- [phosphorus, aromatic attach]
154. -O- [2 phosphorus attach]
155. -O- [phosphorus attach]
156. -O- [arsenic attach]
157. -O- [nitrogen attach]
158. -O- [sulfur attach]
159. -O- [oxygen attach]
160. -O- [aromatic attach]
161. -CH= [aromatic attach]
162. >C= [aromatic attach]
163. -C# [aromatic attach]
164. -CH3 [aromatic attach]
165. -CH2- [aromatic attach]
166. -CH< [aromatic attach]
167. >C< [aromatic attach]
168. -CH3 [aliphatic attach]
169. -CH2- [aliphatic attach]
170. -CH< [aliphatic attach]
171. >C< [aliphatic attach]
172. =C=
173. =CH2 [aliphatic attach]
174. =CH [aliphatic attach]
175. =C [aliphatic attach]
176. #CH [aliphatic attach]
177. #C [aliphatic attach]
178. -NH2 [aromatic attach]

179. -NH- [aromatic attach]
180. -N< [aromatic attach]
181. -NH2 [aliphatic attach]
182. -NH- [aliphatic attach]
183. -N< [aliphatic attach]
184. Fused aromatic carbon
185. Fused aromatic nitrogen
186. ACH
187. ANH
188. AC
189. AN
190. AN+
191. AO
192. AS
193. =O [other]
194. =S [other]
195. -[O-]
196. -O- [aliphatic attach]
197. -S- [aliphatic attach]
198. -F [aromatic attach]
199. -Br [aromatic attach]
200. -Cl [aromatic attach]
201. -I [aromatic attach]
202. -Br [olefinic attach]
203. -Cl [olefinic attach]
204. -F [olefinic attach]
205. -I [olefinic attach]
206. Halogen [Nitrogen attach]
207. -Br [aliphatic attach]

208. -Cl [aliphatic attach]
209. -F [aliphatic attach]
210. -I [aliphatic attach]
211. H [phosphorus attach]
212. H [nitrogen attach]
213. H [silicon attach]
214. H [carbon attach]
215. H [other]
216. -Hg
217. -Hg-
218. -Hg+
219. Hg
220. Pb
221. B
222. P [+3 valence, all single bonds]
223. P [+3 valence, all single bonds]
224. P [+3 valence, all single bonds]
225. P [+5 valence, one double bond]
226. P [+5 valence, two double bonds]
227. P [+5 valence, all single bonds]
228. P
229. As [+3 valence, one double bond]
230. As [+3 valence, all single bonds]
231. As [+5 valence, one double bond]
232. As [+5 valence, two double bonds]
233. As [+5 valence, all single bonds]
234. As

The molecular fragment counts (Martin et al. 2008) represent the number of times various fragments appear in a molecule. The fragments are written in a format similar to SMILES notation (Daylight Chemical Information Systems 2006). The fragments were designed so that each atom in a molecule is only assigned to one fragment. Dashes (-) represent single bonds, Carats (<) represent two single bonds, equal signs (=) represent double bonds, and pound signs (#) represent triple bonds. The terms inside the brackets provide additional information such as what the fragment is attached to. For example,

"-O- [phosphorus, aromatic attach]" is the count of oxygen atoms that are attached to a phosphorus atom and an aromatic atom. The phosphorus atom and the aromatic atom are not considered to be part of the fragment. In addition the term inside the brackets can indicate whether the fragment is part of a ring (for example

"-S- [three membered ring]") or indicate the valence state of a given element (for example "As [+3 valence, one double bond]"). The fragments are assigned in the order that they appear above. For example 1,2-dichlorobenzene would have the following fragment counts:

-Cl [aromatic attach] = 2, AC = 2, and ACH = 4.

VII. 2-D Molecular Properties (9)

1. Number of hydrogen bond acceptors (numHBa)
2. Number of hydrogen bond donors (numHBd)
3. Number of weak hydrogen bond donors (i.e. –CHX, where X = Cl, F) (numwHBd)
4. Sum of E-State indices for hydrogen bond acceptors (SHHBa)
5. Sum of E-State indices for hydrogen bond donors (SHHBd)
6. Sum of E-State indices for weak hydrogen bond donors (SHwHBd)
7. Molecular and group polarity index (Qs)
8. Molecular and group polarity index (Qv)
9. Average of Qs and Qv (Qsv)

The number of hydrogen bond acceptors (numHBa) is the total count of fluorine, oxygen, nonaromatic nitrogens, and aromatic nitrogens with an E-state fragment of SaaN (pyrrolyl nitrogens are not considered to be hydrogen bond acceptors).

The number of hydrogen bond donors (numHBd) is the total count of hydrogens that are attached to oxygen or nitrogen atoms.

The number of weak hydrogen bond donors (numwHBd) is the count of carbon atoms that have both a hydrogen atom and a chlorine or fluorine atom attached.

The sum of E-State indices for hydrogen bond acceptors (SHHBa) is the sum of the E-state indices for atoms that are assigned to be hydrogen bond acceptors.

The sum of E-State indices for hydrogen bond donors (SHHBd) is the sum of the hydrogen E-state indices for oxygens and nitrogens.

The sum of the E-State indices for weak hydrogen bond donors (SHwHBd) is the sum of the hydrogen E-States for the carbon atoms that have both a hydrogen and a chlorine or fluorine attached.

The Qs molecular and group polarity index is given by

$$Q_s = \frac{A^2 \sum_{i=1}^{A} I_i^{alkane}}{\left(\sum_{i=1}^{A} I_i\right)^2} \tag{1}$$

where I_i^{alkane} = The intrinsic state value of an atom considering all non-hydrogen atoms have been replaced by sp3 carbon atoms (keeping the same isoconnectivity).

The Qv molecular and group polarity index is given by

$$Q_v = \frac{\sum_{i=1}^{A} I_i^{max} \sum_{i=1}^{A} I_i^{alkane}}{\left(\sum_{i=1}^{A} I_i\right)^2} \tag{2}$$

where I_i^{max} = The intrinsic state value of the atom where the following replacements have been made:

  (a) all terminal atoms (atoms bonded to one other non-hydrogen atom) replaced by
      –F (I = 8.000)
  (b) all divalent atoms (atoms bonded to two other non-hydrogen atoms) replaced by
      –O- (I = 3.500)
  (c) all trivalent atoms (atoms bonded to three other non-hydrogen atoms) replaced by
      >N- (I = 2.000)
  (d) all quaternary atoms (atoms bonded to four other non-hydrogen atoms) replaced by
      >C< (I = 1.250).

The average of Qs and Qv is given by

$$Q_{sv} = \frac{\sum_{i=1}^{A} I_i^{ave} \sum_{i=1}^{A} I_i^{alkane}}{\left(\sum_{i=1}^{A} I_i \right)^2} \quad (3)$$

where

$$\sum_{i=1}^{A} I_i^{ave} = \frac{\sum_{i=1}^{A} I_i^{alkane} + \sum_{i=1}^{A} I_i^{max}}{2} \quad (4)$$

The equations for Qs, Qv, and Qsv were taken from documentation for MDL QSAR (Elsevier_MDL 2006).

VIII. Information Indices (36)

A. Connectivity Based (8)

1. Information bond index (ib)

2. Mean information content on the edge magnitude (eim)

3. Information content (ic)

4. Mean information content or Shannon's entropy (si)

5. Total information content (I)
6. Maximum information content ($^{max}I_C$ or maxic)
7. Standardized Shannon Information or standardized information content (ssi)
8. Brillouin redundancy index (R)

The information bond index is given by

$$ib = B \log_2(B) - \sum_{g=1}^{G_B} B_g \log_2(B_g) \tag{1}$$

where B = the number of bonds in the molecule, B_g = the number of bonds of type g in the molecule, and G_B = all of the different bond types in the molecule.

The mean information content on the edge magnitude (eim) is based on the magnitude of edge connectivities and is given by

$$eim = -\sum_{b=1}^{B} \frac{(\delta_i \delta_j)_b^{-1/2}}{^1\chi} \log_2 \left(\frac{(\delta_i \delta_j)_b^{-1/2}}{^1\chi} \right) \tag{2}$$

where $^1\chi$ is the first order Chi connectivity index (x1 in Section I), $(\delta_i \delta_j)_b^{-1/2}$ is the edge connectivity of the b bond, and B is the number of bonds.

The information content (a measure of molecular symmetry) of a system having n elements is given by

$$ic = \sum_{g=1}^{G} n_g \log_2(n_g) \tag{3}$$

where n_g is the number of topologically equivalent atoms of type g and G is the number of topologically different atom types. The information content is the same as the zeroth order kappa shape index or Kier symmetry index (Kier 1987). Each equivalence type is constituted by all atoms having the same valence topological state (*Sv*). The valence topological state is given by

$$Sv_i = \sum_{i-1}^{A} t_{ij}^v \tag{4}$$

where t_{ij}^v is the valence topological state matrix given by

$$t_{ij}^v = GM_{ij}^v(1/n_{ij}) \tag{5}$$

where n_{ij} is the number of atoms in the path of mth order. GM_{ij}^v is the geometric mean of the valence vertex degree of the atoms involved in the path between atoms i and j of length m and is defined as:

$$GM_{ij}^v = \left(\prod_{a=1}^{n_{ij}} \delta_a^v \right)^{1/n_{ij}} \tag{6}$$

The mean information content is defined as follows:

$$si = -\sum_{g=1}^{G} p_g \log_2 (p_g) \tag{7}$$

where p_g = the probability (n_g/n) of selecting the gth topological equivalent type. For molecules, the total number of elements in the system (n) is the same as A (the number of non-hydrogen atoms).

The total information content (I) is obtained by multiplying the mean information content by the number of elements:

$$I = n \cdot si \tag{8}$$

The maximum information content (maxic) represents the information content needed to characterize all of the n alternatives:

$$maxic = n \log_2 (n) \tag{9}$$

The standardized information content (ssi) is the ratio between the total information content and the maximum information content:

$$ssi = \frac{I}{maxic} \tag{10}$$

Brillouin redundancy index (R), a measure of the information redundancy of a system is given by
$$R = 1 - ssi \tag{11}$$

The connectivity based information indices are described in more detail on pages 238-240 of the Handbook of Molecular Descriptors (Todeschini and Consonni 2000).

Topologically Based (28)

1. Total information content on the adjacency equality ($^V I_{adj}^E$ or iadje)

2. Mean information content on the adjacency equality ($^V \bar{I}_{adj}^E$ or iadjem)

3. Total information content on the adjacency magnitude ($^V I_{adj}^M$ or iadjm)

4. Mean information content on the adjacency magnitude ($^V \bar{I}_{adj}^M$ or iadjmm)

5. Mean information content of the vertex degree equality ($^V \bar{I}_{adj,deg}^E$ or ivdem)

6. Mean information content of the vertex degree magnitude ($^V \bar{I}_{adj,deg}^M$ or ivdmm)

7. Total information content on the edge adjacency equality ($^E I_{adj}^E$ or ieadje)

8. Mean information content on the edge adjacency equality ($^E \bar{I}_{adj}^E$ or ieadjem)

9. Total information content on the edge adjacency magnitude ($^{E}I_{adj}^{M}$ or ieadjm)

10. Mean information content on the edge adjacency magnitude ($^{E}\bar{I}_{adj}^{M}$ or ieadjmm)

11. Total information content on the distance equality ($^{V}I_{D}^{E}$ or ide)

12. Mean information content on the distance equality ($^{V}\bar{I}_{D}^{E}$ or idem)

13. Total information content on the distance magnitude ($^{V}I_{D}^{M}$ or idm)

14. Mean information content on the distance magnitude ($^{V}\bar{I}_{D}^{M}$ or idmm)

15. Mean information content on the distance degree equality ($^{V}\bar{I}_{D,\text{deg}}^{E}$ or iddem)

16. Mean information content on the distance degree magnitude ($^{V}\bar{I}_{D,\text{deg}}^{M}$ or iddmm)

17. Total information content on the edge distance equality ($^{E}I_{D}^{E}$ or iede)

18. Mean information content on the edge distance equality ($^{E}\bar{I}_{D}^{E}$ or iedem)
19. Edge wiener index (We)
20. Total information content on the edge distance magnitude ($^{E}I_{D}^{M}$ or iedm)
21. Mean information content on the edge distance magnitude ($^{E}\bar{I}_{D}^{M}$ or iedmm)
22. Total vertex cyclicity (tvc)
23. Total information on the vertex cycle matrix equality ($^{V}I_{cyc}^{E}$ or icyce)

24. Mean information on the vertex cycle matrix equality ($^{V}\bar{I}_{cyc}^{E}$ or icycem)

25. Total information on the vertex cycle matrix magnitude ($^{V}I_{cyc}^{M}$ or icycm)

26. Mean information on the vertex cycle matrix magnitude ($^{V}\bar{I}_{cyc}^{M}$ or icycmm)

27. Total topological index (tti)
28. Total topological valence index (ttvi)

The total information content on the adjacency equality (iadje) is given by
$$iadje = A^2 \log_2(A^2) - 2B \log_2(2B) - (A^2 - 2B)\log_2(A^2 - 2B) \qquad (12)$$

The mean information content on the adjacency equality (iadjem) is given by
$$iadjem = -\frac{2B}{A^2}\log_2\left(\frac{2B}{A^2}\right) - \left(1 - \frac{2B}{A^2}\right)\log_2\left(1 - \frac{2B}{A^2}\right) \qquad (13)$$
where A = the number of non-hydrogen atoms in molecule (same as above) and B = the number of bonds in the molecule (same as above). The elements of the Adjacency Matrix, a_{ij}, equal 1 if the ith atom and the jth atom are bonded together and equal 0 if not.

The total information content on the adjacency magnitude (iadjm) is given by
$$iadjm = 2B \log_2(2B) \qquad (14)$$

The mean information content on the adjacency magnitude (iadjmm) is given by
$$iadjmm = 1 + \log_2(B) \qquad (15)$$

The mean information content of the vertex degree equality (ivdem) is given by

$$ivdem = -\sum_{g=1}^{G} \frac{{}^{g}F}{A} \log_2\left(\frac{{}^{g}F}{A}\right) \qquad (16)$$

${}^{g}F$ is the number of vertices with vertex degree equal to g.

The mean information content of the vertex degree magnitude (ivdmm) is given by

$$ivdmm = -\sum_{g=1}^{G} {}^{g}F \frac{g}{2B} \log_2\left(\frac{g}{2B}\right) \qquad (17)$$

The total information content on the edge adjacency equality (ieadje) is given by

$$ieadje = B^2 \log_2(B^2) - 2N_2 \log_2(2N_2) - (B^2 - 2N_2)\log_2(B^2 - 2N_2) \qquad (18)$$

The mean information content on the edge adjacency equality (ieadjem) is given by

$$ieadjem = -\frac{2N_2}{B^2} \log_2\left(\frac{2N_2}{B^2}\right) - \left(1 - \frac{2N_2}{B^2}\right)\log_2\left(1 - \frac{2N_2}{B^2}\right) \qquad (19)$$

where N_2 = the number of 2nd order path counts in the molecule.

The total information content on the edge adjacency magnitude (ieadjm) is given by

$$ieadjm = 2N_2 \log_2(2N_2) \qquad (20)$$

The mean information content on the edge adjacency magnitude (ieadjmm) is given by

$$ieadjmm = 1 + \log_2(N_2) \qquad (21)$$

The total information content on the distance equality (ide) is given by

$$ide = \frac{A(A-1)}{2} \log_2\left(\frac{A(A-1)}{2}\right) - \sum_{g=1}^{G} {}^{g}f \log_2({}^{g}f) \qquad (22)$$

where ${}^{g}f$ is the number of distances with equal g values in the triangular D submatrix. D is an A x A matrix that contains the graph distances between atoms. The graph distances are calculated as $1/$(the # of bonds between atoms)^2.

The mean information content on the distance equality (idem) is given by

$$idem = -\sum_{g=1}^{G} \frac{2 \cdot {}^{g}f}{A(A-1)} \log_2\left(\frac{2 \cdot {}^{g}f}{A(A-1)}\right) \qquad (23)$$

The total information content on the distance magnitude (idm) is given by

$$idm = W \log_2(W) - \sum {}^{g}f \cdot g \log_2(g) \qquad (24)$$

The mean information content on the distance magnitude (idmm) is given by

$$idmm = -\sum_{g=1}^{G} {}^{g}f \cdot \frac{g}{W} \log_2\left(\frac{g}{W}\right) \qquad (25)$$

The mean information content on the distance degree equality (iddem)

$$iddem = -\sum_{g-1}^{G} \frac{n_g}{A} \log_2\left(\frac{n_g}{A}\right) \tag{26}$$

where n_g is the number of vertices having equal vertex distance degrees (σ) in the gth class, G is the number of equivalence classes.

The mean information content on the distance degree magnitude (iddmm) is given by

$$iddmm = -\sum_{g-1}^{G} n_g \frac{\sigma_g}{2W} \log_2\left(\frac{\sigma_g}{2W}\right) \tag{27}$$

The total information content on the edge distance equality (iede) is given by

$$iede = \frac{B(B-1)}{2} \log_2\left(\frac{B(B-1)}{2}\right) - \sum {}^g f \log_2\left({}^g f\right) \tag{28}$$

The mean information content on the edge distance equality (iedem) is given by

$$iedem = -\sum \frac{2 \cdot {}^g f}{B(B-1)} \log_2\left(\frac{2 \cdot {}^g f}{B(B-1)}\right) \tag{29}$$

The edge wiener index (We) is given by

$$We = W - \frac{A(A-1)}{2} \tag{30}$$

The total information content on the edge distance magnitude (iedm) is given by

$$iedm = We \log_2(We) - \sum {}^g f \cdot g \log_2(g) \tag{31}$$

The mean information content on the edge distance magnitude (iedmm) is given by

$$iedmm = -\sum {}^g f \cdot \frac{g}{We} \log_2\left(\frac{g}{We}\right) \tag{32}$$

The total vertex cyclicity (tvc) is given by

$$tvc = \sum_{i=1}^{A} \sum_{j=1}^{C+} c_{ij} \tag{33}$$

where C+ = the number of cycles in a molecule (e.g. Naphthalene has 3 cycles) and $c_{ij} = 1$ if the ith atom is part of the jth cycle and 0 otherwise. The elements c_{ij} comprise the vertex cycle matrix C_V.

The total information on the vertex cycle matrix equality (icyce) is given by

$$icyce = A \cdot C^+ \log_2\left(A \cdot C^+\right) - n_1 \log_2(n_1) - n_0 \log_2(n_0) \tag{34}$$

where n_0 = the total number of c_{ij} elements that are equal to 0 and n_1 = the total number of c_{ij} elements that are equal to 1.

The mean information on the vertex cycle matrix equality (icycem) is given by

$$icycem = -\frac{n_1}{A \cdot C^+} \log_2\left(\frac{n_1}{A \cdot C^+}\right) - \frac{n_0}{A \cdot C^+} \log_2\left(\frac{n_0}{A \cdot C^+}\right) \tag{35}$$

The total information on the vertex cycle matrix magnitude (icycm) is given by
$$icycm = n_1 \log_2(n_1) \tag{36}$$

The mean information on the vertex cycle matrix magnitude (icycmm) is given by
$$icycmm = \log_2(n_1) \tag{37}$$

The total topological index (tti) is given by
$$tti = \sum_{i=1}^{A} t_{ii} + \sum_{i=1}^{A-1} \sum_{j=i+1}^{A} t_{ij} \tag{38}$$

where
$$t_{ij} = GM_{ij} \cdot (1/n_{ij}) \tag{39}$$

where GM_{ij} is the geometric mean of the vertex degree of the atoms involved in the path i-j of length m:

$$GM_{ij} = \left(\prod_{a=1}^{n_{ij}} \delta_a\right)^{1/n_{ij}} \tag{40}$$

where δ_a is the vertex degree of atom a (defined in Section II) and n_{ij} = the number of atoms in the path from atom i to atom j.

The total topological valence index (ttvi) is given by

$$ttiv = \sum_{i=1}^{A} t_{ii}^{v} + \sum_{i=1}^{A-1} \sum_{j=i+1}^{A} t_{ij}^{v} \tag{41}$$

where the elements of the valence topological state matrix are calculated using equations 5 and 6. The topologically based information indices are described in greater detail on pages 447-457 of the Handbook of Molecular Descriptors (Todeschini and Consonni 2000).

IX. Molecular Distance-Edge Vector (19)

1. Molecular distance edge between all primary carbons (MDEC11)
2. Molecular distance edge between all primary and secondary carbons (MDEC12)
3. Molecular distance edge between all primary and tertiary carbons (MDEC13)
4. Molecular distance edge between all primary and quaternary carbons (MDEC14)
5. Molecular distance edge between all secondary carbons (MDEC22)
6. Molecular distance edge between all secondary and tertiary carbons (MDEC23)
7. Molecular distance edge between all secondary and quaternary carbons (MDEC24)
8. Molecular distance edge between all tertiary carbons (MDEC33)
9. Molecular distance edge between all tertiary and quaternary carbons (MDEC34)
10. Molecular distance edge between all quaternary carbons (MDEC44)
11. Molecular distance edge between all primary oxygens (MDEO11)
12. Molecular distance edge between all primary and secondary oxygens (MDEO12)
13. Molecular distance edge between all secondary oxygens (MDEO22)
14. Molecular distance edge between all primary nitrogens (MDEN11)
15. Molecular distance edge between all primary and secondary nitrogens (MDEN12)
16. Molecular distance edge between all primary and tertiary nitrogens (MDEN13)
17. Molecular distance edge between all secondary nitrogens (MDEN22)
18. Molecular distance edge between all secondary and tertiary nitrogens (MDEN23)
19. Molecular distance edge between all tertiary nitrogens (MDEN33)

The general formula for the molecule distance edge vector descriptors is as follows:

$$\lambda_{st} = \frac{n_{st}}{\bar{d}_{st}^{\,2}} \tag{1}$$

where λ_{st} is the molecular distance-edge vector between atom types s and t, n_{st} is the number of possible atom pairs for a fixed combination of s and t, and $\bar{d}_{st}$ is given by

$$\bar{d}_{st} = \prod_{s \leq t} \left(d_{i(s),j(t)} \right)^{1/(2 \cdot n_{st})} \tag{2}$$

where $d_{i(s),j(t)}$ is the distance (the shortest path length) between atom i with types s and atom j with type t. The geometric mean takes into account all the distances between carbon atoms i and j of types s and t. The descriptors are denoted as MDEXst where X is the element, s is the first atom type and t is the second atom type. The atom type or degree (primary, secondary, or tertiary) is determined from the count of non-hydrogen atoms that are attached to a given atom.

The molecular distance edge vectors are described on page 116 of the Handbook of Molecular Descriptors (Todeschini and Consonni 2000).

X. Burden Eigenvalue Descriptors (64)

1. Highest eigenvalue n. 1 of Burden matrix / weighted by atomic masses (BEHm1)
2. Highest eigenvalue n. 2 of Burden matrix / weighted by atomic masses (BEHm2)
3. Highest eigenvalue n. 3 of Burden matrix / weighted by atomic masses (BEHm3)
4. Highest eigenvalue n. 4 of Burden matrix / weighted by atomic masses (BEHm4)
5. Highest eigenvalue n. 5 of Burden matrix / weighted by atomic masses (BEHm5)
6. Highest eigenvalue n. 6 of Burden matrix / weighted by atomic masses (BEHm6)
7. Highest eigenvalue n. 7 of Burden matrix / weighted by atomic masses (BEHm7)
8. Highest eigenvalue n. 8 of Burden matrix / weighted by atomic masses (BEHm8)
9. Lowest eigenvalue n. 1 of Burden matrix / weighted by atomic masses (BELm1)
10. Lowest eigenvalue n. 2 of Burden matrix / weighted by atomic masses (BELm2)
11. Lowest eigenvalue n. 3 of Burden matrix / weighted by atomic masses (BELm3)
12. Lowest eigenvalue n. 4 of Burden matrix / weighted by atomic masses (BELm4)
13. Lowest eigenvalue n. 5 of Burden matrix / weighted by atomic masses (BELm5)
14. Lowest eigenvalue n. 6 of Burden matrix / weighted by atomic masses (BELm6)
15. Lowest eigenvalue n. 7 of Burden matrix / weighted by atomic masses (BELm7)
16. Lowest eigenvalue n. 8 of Burden matrix / weighted by atomic masses (BELm8)
17. Highest eigenvalue n. 1 of Burden matrix / weighted by atomic van der Waals volumes (BEHv1)
18. Highest eigenvalue n. 2 of Burden matrix / weighted by atomic van der Waals volumes (BEHv2)
19. Highest eigenvalue n. 3 of Burden matrix / weighted by atomic van der Waals volumes (BEHv3)
20. Highest eigenvalue n. 4 of Burden matrix / weighted by atomic van der Waals volumes (BEHv4)
21. Highest eigenvalue n. 5 of Burden matrix / weighted by atomic van der Waals volumes (BEHv5)
22. Highest eigenvalue n. 6 of Burden matrix / weighted by atomic van der Waals volumes (BEHv6)
23. Highest eigenvalue n. 7 of Burden matrix / weighted by atomic van der Waals volumes (BEHv7)

24. Highest eigenvalue n. 8 of Burden matrix / weighted by atomic van der Waals volumes (BEHv8)

25. Lowest eigenvalue n. 1 of Burden matrix / weighted by atomic van der Waals volumes (BELv1)

26. Lowest eigenvalue n. 2 of Burden matrix / weighted by atomic van der Waals volumes (BELv2)

27. Lowest eigenvalue n. 3 of Burden matrix / weighted by atomic van der Waals volumes (BELv3)

28. Lowest eigenvalue n. 4 of Burden matrix / weighted by atomic van der Waals volumes (BELv4)

29. Lowest eigenvalue n. 5 of Burden matrix / weighted by atomic van der Waals volumes (BELv5)

30. Lowest eigenvalue n. 6 of Burden matrix / weighted by atomic van der Waals volumes (BELv6)

31. Lowest eigenvalue n. 7 of Burden matrix / weighted by atomic van der Waals volumes (BELv7)

32. Lowest eigenvalue n. 8 of Burden matrix / weighted by atomic van der Waals volumes (BELv8)

33. Highest eigenvalue n. 1 of Burden matrix / weighted by atomic Sanderson electronegativities (BEHe1)

34. Highest eigenvalue n. 2 of Burden matrix / weighted by atomic Sanderson electronegativities (BEHe2)

35. Highest eigenvalue n. 3 of Burden matrix / weighted by atomic Sanderson electronegativities (BEHe3)

36. Highest eigenvalue n. 4 of Burden matrix / weighted by atomic Sanderson electronegativities (BEHe4)

37. Highest eigenvalue n. 5 of Burden matrix / weighted by atomic Sanderson electronegativities (BEHe5)

38. Highest eigenvalue n. 6 of Burden matrix / weighted by atomic Sanderson electronegativities (BEHe6)

39. Highest eigenvalue n. 7 of Burden matrix / weighted by atomic Sanderson electronegativities (BEHe7)

40. Highest eigenvalue n. 8 of Burden matrix / weighted by atomic Sanderson electronegativities (BEHe8)

41. Lowest eigenvalue n. 1 of Burden matrix / weighted by atomic Sanderson electronegativities (BELe1)

42. Lowest eigenvalue n. 2 of Burden matrix / weighted by atomic Sanderson electronegativities (BELe2)

43. Lowest eigenvalue n. 3 of Burden matrix / weighted by atomic Sanderson electronegativities (BELe3)

44. Lowest eigenvalue n. 4 of Burden matrix / weighted by atomic Sanderson electronegativities (BELe4)

45. Lowest eigenvalue n. 5 of Burden matrix / weighted by atomic Sanderson electronegativities (BELe5)

46. Lowest eigenvalue n. 6 of Burden matrix / weighted by atomic Sanderson electronegativities (BELe6)

47. Lowest eigenvalue n. 7 of Burden matrix / weighted by atomic Sanderson electronegativities (BELe7)

48. Lowest eigenvalue n. 8 of Burden matrix / weighted by atomic Sanderson electronegativities (BELe8)

49. Highest eigenvalue n. 1 of Burden matrix / weighted by atomic polarizabilities (BEHp1)

50. Highest eigenvalue n. 2 of Burden matrix / weighted by atomic polarizabilities (BEHp2)

51. Highest eigenvalue n. 3 of Burden matrix / weighted by atomic polarizabilities (BEHp3)

52. Highest eigenvalue n. 4 of Burden matrix / weighted by atomic polarizabilities (BEHp4)

53. Highest eigenvalue n. 5 of Burden matrix / weighted by atomic polarizabilities (BEHp5)

54. Highest eigenvalue n. 6 of Burden matrix / weighted by atomic polarizabilities (BEHp6)

55. Highest eigenvalue n. 7 of Burden matrix / weighted by atomic polarizabilities (BEHp7)

56. Highest eigenvalue n. 8 of Burden matrix / weighted by atomic polarizabilities (BEHp8)

57. Lowest eigenvalue n. 1 of Burden matrix / weighted by atomic polarizabilities (BELp1)

58. Lowest eigenvalue n. 2 of Burden matrix / weighted by atomic polarizabilities (BELp2)

59. Lowest eigenvalue n. 3 of Burden matrix / weighted by atomic polarizabilities (BELp3)

60. Lowest eigenvalue n. 4 of Burden matrix / weighted by atomic polarizabilities (BELp4)

61. Lowest eigenvalue n. 5 of Burden matrix / weighted by atomic polarizabilities (BELp5)

62. Lowest eigenvalue n. 6 of Burden matrix / weighted by atomic polarizabilities (BELp6)

63. Lowest eigenvalue n. 7 of Burden matrix / weighted by atomic polarizabilities (BELp7)

64. Lowest eigenvalue n. 8 of Burden matrix / weighted by atomic polarizabilities (BELp8)

The Burden eigenvalue descriptors are determined by solving the following general eigenvalue equation:

$$\boldsymbol{B} \cdot \boldsymbol{V} = \boldsymbol{V} \cdot \boldsymbol{e} \qquad (1)$$

where $\boldsymbol{B}$ is a real connectivity matrix to be defined, $\boldsymbol{V}$ is a matrix of eigenvectors, and $\boldsymbol{e}$ is a diagonal matrix of eigenvalues. The rules defining $\boldsymbol{B}$ are as follows:

a. Hydrogen atoms *are* included.

b. The diagonal elements of $\boldsymbol{B}$, B_{ii}, are either given by the carbon normalized atomic mass, van der Waals volume, Sanderson electronegativity, and polarizability of atom i.

c. The element of $\boldsymbol{B}$ connecting atoms i and j, B_{ij}, is equal to the square root of the bond order between atoms i and j.

d. All other elements of B (corresponding non bonded atom pairs) are set to 0.001.

The carbon normalized weights used in the software as follows:

Symbol	Mass	VdW Volume	Electronegativity	Polarizability
C	1.000	1.000	1.000	1.000
H	0.084	0.299	0.942	0.381
N	1.167	0.695	1.160	0.625
O	1.332	0.512	1.327	0.455
P	2.579	1.181	0.916	2.063
S	2.670	1.088	1.076	1.648
F	1.582	0.410	1.455	0.318
Cl	2.952	1.000	1.265	1.239
Br	6.653	1.384	1.171	1.733
I	10.566	1.728	1.011	3.040
Si	2.339	1.424	0.778	3.057
As	6.238	1.181	1.025	2.449
Sn	9.884	2.041	0.836	4.375
Hg	16.702	0.695	0.800	3.239
Pb	17.252	1.538	0.833	3.864

The lowest eigenvalues are the absolute values of the negative eigenvalues. The highest eigenvalues are the eight largest positive eigenvalues. A routine for solving for eigenvalues is found in the Jama Java package (National_Institute_of_Standards_and_Technology 2005). The Burden eigenvalues descriptors are described on pages 132-133 of the Handbook of Molecular Descriptors(Todeschini and Consonni 2000).

XI. Walk and Path Counts (34)

1. Molecular walk count of order 01 (MWC01)
2. Molecular walk count of order 02 (MWC02)
3. Molecular walk count of order 03 (MWC03)
4. Molecular walk count of order 04 (MWC04)
5. Molecular walk count of order 05 (MWC05)
6. Molecular walk count of order 06 (MWC06)
7. Molecular walk count of order 07 (MWC07)
8. Molecular walk count of order 08 (MWC08)
9. Molecular walk count of order 09 (MWC09)
10. Molecular walk count of order 10 (MWC10)
11. Total walk count(TWC)
12. Self-returning walk count of order 01 (SRW01)
13. Self-returning walk count of order 02 (SRW02)
14. Self-returning walk count of order 03 (SRW03)
15. Self-returning walk count of order 04 (SRW04)
16. Self-returning walk count of order 05 (SRW05)
17. Self-returning walk count of order 06 (SRW06)
18. Self-returning walk count of order 07 (SRW07)
19. Self-returning walk count of order 08 (SRW08)
20. Self-returning walk count of order 09 (SRW09)
21. Self-returning walk count of order 10 (SRW10)
22. Molecular path count of order 01 (MPC01)
23. Molecular path count of order 02 (MPC02)
24. Molecular path count of order 03 (MPC03)
25. Molecular path count of order 04 (MPC04)
26. Molecular path count of order 05 (MPC05)
27. Molecular path count of order 06 (MPC06)
28. Molecular path count of order 07 (MPC07)
29. Molecular path count of order 08 (MPC08)
30. Molecular path count of order 09 (MPC09)
31. Molecular path count of order 10 (MPC10)
32. Total path count(TPC)
33. Molecular multiple path count of order 01 (piPC01)
34. Molecular multiple path count of order 02 (piPC02)
35. Molecular multiple path count of order 03 (piPC03)
36. Molecular multiple path count of order 04 (piPC04)
37. Molecular multiple path count of order 05 (piPC05)
38. Molecular multiple path count of order 06 (piPC06)
39. Molecular multiple path count of order 07 (piPC07)
40. Molecular multiple path count of order 08 (piPC08)
41. Molecular multiple path count of order 09 (piPC09)
42. Molecular multiple path count of order 10 (piPC10)
43. Conventional bond-order ID number(piID)
44. Randic ID number(CID)

45. Average Randic ID number(CID2)
46. Balaban ID number(BID)

The molecular walk count of kth order (MWCk) is the total number of walks of the kth length in the hydrogen suppressed molecular graph given by

$$MWCk = \frac{1}{2} \cdot \sum_{i=1}^{A} \sum_{j=1}^{A} a_{ij}^{(k)}$$
(1)

where $a_{ij}^{(k)}$ are the elements of A^k (the kth power of the adjacency matrix A) and A is the number of atoms. To match Dragon (Talete 2006) the following formula was used if k is greater than one:

$$MWCk = \ln\left(\sum_{i=1}^{A} \sum_{j=1}^{A} a_{ij}^{(k)} + 1\right)$$
(2)

The molecular walk count descriptors are described on page 481 of the Handbook of Molecular Descriptors(Todeschini and Consonni 2000).

The total walk count (TWC) is given by

$$TWC = \sum_{k=0}^{A-1} MWCk = A + \sum_{k=1}^{A-1} MWCk$$
(3)

To match Dragon the following formula must be used:

$$TWC = A + \sum_{k=1}^{10} MWCk$$
(4)

The self-returning walk count of kth order (SRWk) is the total number of self-returning walks of length k in the graph and calculated as follows:

$$SRWk = tr(A^k)$$
(5)

where tr is the trace operator (sum of the diagonal elements) and A^k is the kth power of the adjacency matrix. The self-returning walk count descriptors are described on page 384 of the Handbook of Molecular Descriptors(Todeschini and Consonni 2000).

The molecular path counts (MPCk) are simply the number of unique paths of length k. The molecular path count descriptors are described on page 345 of the Handbook of Molecular Descriptors(Todeschini and Consonni 2000).

The total path count (TPC) is obtained by summing all the molecular path counts:

$$TPC = \sum_{i=0}^{L} MPCk = A + \sum_{k=1}^{L} MPCk$$
(6)

where L is the maximum path length. In order to match Dragon the total path count is instead given by

$$TPC = A + \sum_{k=1}^{L} \ln(1 + MPCk)$$
(7)

The molecular multiple path counts (piPCk) are defined as path counts weighted by the bond order:

$$piPCk = \sum_{{}^k p_{ij}} w_{ij} \tag{8}$$

where ${}^k p_{ij}$ denotes a path of length k from vertex i to vertex j and w_{ij} is the path weight given by the product of the bond orders in the path:

$$w_{ij} = \prod_{b=1}^{k} \pi_b^* \tag{9}$$

It should be noted that aromatic bonds are assigned a bond order of 1.5. In order to match Dragon, again a similar logarithmic transformation is used to calculate *piPCk*:

$$piPCk = \ln\left(1 + \sum_{{}^k p_{ij}} w_{ij}\right) \tag{10}$$

The conventional bond-order ID number(piID) is calculated as the sum of the molecular multiple path counts:

$$piID = \sum_{k=0}^{L} piPCk = A + \sum_{k=1}^{L} piPCk \tag{11}$$

The Randic ID number is also defined as a weighted molecular path count:

$$CID = A + \sum_{{}^k p_{ij}} w_{ij} \tag{12}$$

where A is the number of atoms, ${}^k p_{ij}$ denotes a path of length k from vertex i to vertex j and w_{ij} is the path weight given by

$$w_{ij} = \prod_{b=1}^{m} \left(\delta_{b(1)} \cdot \delta_{b(2)}\right)_b^{-1/2} \tag{13}$$

where $\delta_{b(1)}$ and $\delta_{b(2)}$ are the vertex degree of the two atoms incident to the *bth* edge and b runs over all of the *m* edges of the path.

The average Randic ID number is given by

$$CID2 = CID / A \tag{14}$$

The Balaban ID number *(BID)* is calculated using the same formula as Randic ID number (*CID*) except the path weight is given by

$$w_{ij} = \prod_{b=1}^{m} \left(\sigma_{b(1)} \cdot \sigma_{b(2)}\right)_b^{-1/2} \tag{15}$$

where $\sigma_{b(1)}$ and $\sigma_{b(2)}$ are the vertex distance degrees of the two atoms incident to the *bth* edge and b runs over all of the *m* edges of the path.

CID and BID are described on pages 227 and 229, respectively of the Handbook of Molecular Descriptors (Todeschini and Consonni 2000).

XII. 2D Autocorrelation Descriptors

1. Broto-Moreau autocorrelation of a topological structure - lag 1 / weighted by atomic masses (ATS1m)
2. Broto-Moreau autocorrelation of a topological structure - lag 2 / weighted by atomic masses (ATS2m)
3. Broto-Moreau autocorrelation of a topological structure - lag 3 / weighted by atomic masses (ATS3m)
4. Broto-Moreau autocorrelation of a topological structure - lag 4 / weighted by atomic masses (ATS4m)
5. Broto-Moreau autocorrelation of a topological structure - lag 5 / weighted by atomic masses (ATS5m)
6. Broto-Moreau autocorrelation of a topological structure - lag 6 / weighted by atomic masses (ATS6m)
7. Broto-Moreau autocorrelation of a topological structure - lag 7 / weighted by atomic masses (ATS7m)
8. Broto-Moreau autocorrelation of a topological structure - lag 8 / weighted by atomic masses (ATS8m)
9. Broto-Moreau autocorrelation of a topological structure - lag 1 / weighted by atomic van der Waals volumes (ATS1v)
10. Broto-Moreau autocorrelation of a topological structure - lag 2 / weighted by atomic van der Waals volumes (ATS2v)
11. Broto-Moreau autocorrelation of a topological structure - lag 3 / weighted by atomic van der Waals volumes (ATS3v)
12. Broto-Moreau autocorrelation of a topological structure - lag 4 / weighted by atomic van der Waals volumes (ATS4v)
13. Broto-Moreau autocorrelation of a topological structure - lag 5 / weighted by atomic van der Waals volumes (ATS5v)
14. Broto-Moreau autocorrelation of a topological structure - lag 6 / weighted by atomic van der Waals volumes (ATS6v)
15. Broto-Moreau autocorrelation of a topological structure - lag 7 / weighted by atomic van der Waals volumes (ATS7v)
16. Broto-Moreau autocorrelation of a topological structure - lag 8 / weighted by atomic van der Waals volumes (ATS8v)
17. Broto-Moreau autocorrelation of a topological structure - lag 1 / weighted by atomic Sanderson electronegativities (ATS1e)
18. Broto-Moreau autocorrelation of a topological structure - lag 2 / weighted by atomic Sanderson electronegativities (ATS2e)
19. Broto-Moreau autocorrelation of a topological structure - lag 3 / weighted by atomic Sanderson electronegativities (ATS3e)
20. Broto-Moreau autocorrelation of a topological structure - lag 4 / weighted by atomic Sanderson electronegativities (ATS4e)
21. Broto-Moreau autocorrelation of a topological structure - lag 5 / weighted by atomic Sanderson electronegativities (ATS5e)
22. Broto-Moreau autocorrelation of a topological structure - lag 6 / weighted by atomic Sanderson electronegativities (ATS6e)

23. Broto-Moreau autocorrelation of a topological structure - lag 7 / weighted by atomic Sanderson electronegativities (ATS7e)
24. Broto-Moreau autocorrelation of a topological structure - lag 8 / weighted by atomic Sanderson electronegativities (ATS8e)
25. Broto-Moreau autocorrelation of a topological structure - lag 1 / weighted by atomic polarizabilities (ATS1p)
26. Broto-Moreau autocorrelation of a topological structure - lag 2 / weighted by atomic polarizabilities (ATS2p)
27. Broto-Moreau autocorrelation of a topological structure - lag 3 / weighted by atomic polarizabilities (ATS3p)
28. Broto-Moreau autocorrelation of a topological structure - lag 4 / weighted by atomic polarizabilities (ATS4p)
29. Broto-Moreau autocorrelation of a topological structure - lag 5 / weighted by atomic polarizabilities (ATS5p)
30. Broto-Moreau autocorrelation of a topological structure - lag 6 / weighted by atomic polarizabilities (ATS6p)
31. Broto-Moreau autocorrelation of a topological structure - lag 7 / weighted by atomic polarizabilities (ATS7p)
32. Broto-Moreau autocorrelation of a topological structure - lag 8 / weighted by atomic polarizabilities (ATS8p)
33. Moran autocorrelation - lag 1 / weighted by atomic masses (MATS1m)
34. Moran autocorrelation - lag 2 / weighted by atomic masses (MATS2m)
35. Moran autocorrelation - lag 3 / weighted by atomic masses (MATS3m)
36. Moran autocorrelation - lag 4 / weighted by atomic masses (MATS4m)
37. Moran autocorrelation - lag 5 / weighted by atomic masses (MATS5m)
38. Moran autocorrelation - lag 6 / weighted by atomic masses (MATS6m)
39. Moran autocorrelation - lag 7 / weighted by atomic masses (MATS7m)
40. Moran autocorrelation - lag 8 / weighted by atomic masses (MATS8m)
41. Moran autocorrelation - lag 1 / weighted by atomic van der Waals volumes (MATS1v)
42. Moran autocorrelation - lag 2 / weighted by atomic van der Waals volumes (MATS2v)
43. Moran autocorrelation - lag 3 / weighted by atomic van der Waals volumes (MATS3v)
44. Moran autocorrelation - lag 4 / weighted by atomic van der Waals volumes (MATS4v)
45. Moran autocorrelation - lag 5 / weighted by atomic van der Waals volumes (MATS5v)
46. Moran autocorrelation - lag 6 / weighted by atomic van der Waals volumes (MATS6v)
47. Moran autocorrelation - lag 7 / weighted by atomic van der Waals volumes (MATS7v)
48. Moran autocorrelation - lag 8 / weighted by atomic van der Waals volumes (MATS8v)
49. Moran autocorrelation - lag 1 / weighted by atomic Sanderson electronegativities (MATS1e)
50. Moran autocorrelation - lag 2 / weighted by atomic Sanderson electronegativities (MATS2e)
51. Moran autocorrelation - lag 3 / weighted by atomic Sanderson electronegativities (MATS3e)
52. Moran autocorrelation - lag 4 / weighted by atomic Sanderson electronegativities (MATS4e)
53. Moran autocorrelation - lag 5 / weighted by atomic Sanderson electronegativities (MATS5e)

54. Moran autocorrelation - lag 6 / weighted by atomic Sanderson electronegativities (MATS6e)
55. Moran autocorrelation - lag 7 / weighted by atomic Sanderson electronegativities (MATS7e)
56. Moran autocorrelation - lag 8 / weighted by atomic Sanderson electronegativities (MATS8e)
57. Moran autocorrelation - lag 1 / weighted by atomic polarizabilities (MATS1p)
58. Moran autocorrelation - lag 2 / weighted by atomic polarizabilities (MATS2p)
59. Moran autocorrelation - lag 3 / weighted by atomic polarizabilities (MATS3p)
60. Moran autocorrelation - lag 4 / weighted by atomic polarizabilities (MATS4p)
61. Moran autocorrelation - lag 5 / weighted by atomic polarizabilities (MATS5p)
62. Moran autocorrelation - lag 6 / weighted by atomic polarizabilities (MATS6p)
63. Moran autocorrelation - lag 7 / weighted by atomic polarizabilities (MATS7p)
64. Moran autocorrelation - lag 8 / weighted by atomic polarizabilities (MATS8p)
65. Geary autocorrelation - lag 1 / weighted by atomic masses (GATS1m)
66. Geary autocorrelation - lag 2 / weighted by atomic masses (GATS2m)
67. Geary autocorrelation - lag 3 / weighted by atomic masses (GATS3m)
68. Geary autocorrelation - lag 4 / weighted by atomic masses (GATS4m)
69. Geary autocorrelation - lag 5 / weighted by atomic masses (GATS5m)
70. Geary autocorrelation - lag 6 / weighted by atomic masses (GATS6m)
71. Geary autocorrelation - lag 7 / weighted by atomic masses (GATS7m)
72. Geary autocorrelation - lag 8 / weighted by atomic masses (GATS8m)
73. Geary autocorrelation - lag 1 / weighted by atomic van der Waals volumes (GATS1v)
74. Geary autocorrelation - lag 2 / weighted by atomic van der Waals volumes (GATS2v)
75. Geary autocorrelation - lag 3 / weighted by atomic van der Waals volumes (GATS3v)
76. Geary autocorrelation - lag 4 / weighted by atomic van der Waals volumes (GATS4v)
77. Geary autocorrelation - lag 5 / weighted by atomic van der Waals volumes (GATS5v)
78. Geary autocorrelation - lag 6 / weighted by atomic van der Waals volumes (GATS6v)
79. Geary autocorrelation - lag 7 / weighted by atomic van der Waals volumes (GATS7v)
80. Geary autocorrelation - lag 8 / weighted by atomic van der Waals volumes (GATS8v)
81. Geary autocorrelation - lag 1 / weighted by atomic Sanderson electronegativities (GATS1e)
82. Geary autocorrelation - lag 2 / weighted by atomic Sanderson electronegativities (GATS2e)
83. Geary autocorrelation - lag 3 / weighted by atomic Sanderson electronegativities (GATS3e)
84. Geary autocorrelation - lag 4 / weighted by atomic Sanderson electronegativities (GATS4e)
85. Geary autocorrelation - lag 5 / weighted by atomic Sanderson electronegativities (GATS5e)
86. Geary autocorrelation - lag 6 / weighted by atomic Sanderson electronegativities (GATS6e)
87. Geary autocorrelation - lag 7 / weighted by atomic Sanderson electronegativities (GATS7e)
88. Geary autocorrelation - lag 8 / weighted by atomic Sanderson electronegativities (GATS8e)

89. Geary autocorrelation - lag 1 / weighted by atomic polarizabilities (GATS1p)
90. Geary autocorrelation - lag 2 / weighted by atomic polarizabilities (GATS2p)
91. Geary autocorrelation - lag 3 / weighted by atomic polarizabilities (GATS3p)
92. Geary autocorrelation - lag 4 / weighted by atomic polarizabilities (GATS4p)
93. Geary autocorrelation - lag 5 / weighted by atomic polarizabilities (GATS5p)
94. Geary autocorrelation - lag 6 / weighted by atomic polarizabilities (GATS6p)
95. Geary autocorrelation - lag 7 / weighted by atomic polarizabilities (GATS7p)
96. Geary autocorrelation - lag 8 / weighted by atomic polarizabilities (GATS8p)

The Broto-Moreau autocorrelation descriptors (ATSdw) are given by

$$ATSdw = \sum_{i=1}^{A}\sum_{j=1}^{A} \delta_{ij} \cdot w_i \cdot w_j \tag{1}$$

where d is the considered topological distance (i.e. the lag in the autocorrelation terms), δ_{ij} is the Kronecker delta function ($\delta_{ij} = 1$ if $d_{ij} = d$, zero otherwise), and w_i and w_j are the weights (normalized atomic properties) for atoms i and j respectively. The normalized atomic mass, van der Waals volume, electronegativity, or polarizability can be used for the weights. To match Dragon(Talete 2006), the Broto-Moreau autocorrelation descriptors are calculated in the software as follows:

$$ATSdw = \ln\left(1 + \sum_{i=1}^{A}\sum_{j=1}^{A} \delta_{ij} \cdot w_i \cdot w_j\right) \tag{2}$$

The Moran autocorrelation descriptors (MATSdw) are given by

$$MATSdw = \frac{\frac{1}{\Delta} \cdot \sum_{i-1}^{A}\sum_{j=1}^{A} \delta_{ij} \cdot (w_i - \overline{w}) \cdot (w_j - \overline{w})}{\frac{1}{A}\sum_{i=1}^{A}(w_i - \overline{w})^2} \tag{3}$$

where $\overline{w}$ is the average value of the property for the molecule and
Δ is the number of vertex pairs at distance equal to d.

The Geary autocorrelation descriptors are given by

$$GATSdw = \frac{\frac{1}{2\Delta} \cdot \sum_{i-1}^{A}\sum_{j=1}^{A} \delta_{ij} \cdot \left(w_i - w_j\right)^2}{\frac{1}{A-1}\sum_{i=1}^{A}\left(w_i - \overline{w}\right)^2} \qquad (4)$$

The carbon normalized weights are given in Section X (Burden eigenvalue descriptors).

The 2D autocorrelation descriptors are described on page 17-19 of the Handbook of Molecular Descriptors (Todeschini and Consonni 2000).

XIII. Molecular properties (7)

1. Ghose-Crippen octanol water coefficient (ALOGP)
2. Ghose-Crippen octanol water coefficient squared (ALOGP2)
3. Ghose-Crippen molar refractivity (AMR)
4. Wang octanol water partition coefficient (XLOGP)
5. Wang octanol water partition coefficient squared (XLOGP2)
6. Hydrophilic index (Hy)
7. Unsaturation index (Ui)

The Ghose-Crippen octanol water coefficient (ALOGP) is a group contribution model for the octanol-water partition coefficient. ALOGP is defined as follows:

$$ALOGP = \sum_k a_k N_k \tag{1}$$

where a_k is the group contribution coefficient for the kth fragment type and N_k is the number of occurrences for the kth fragment type. The ALOGP descriptor is described in greater detail on page 275 of the Handbook of Molecular Descriptors (Todeschini and Consonni 2000) and by Ghose and coworkers (Ghose et al. 1998).

ALOGP2 is simply the square of ALOGP:

$$ALOGP2 = ALOGP^2 \tag{2}$$

The Ghose-Crippen molar refractivity (AMR) is calculated using a similar group contribution approach:

$$AMR = \sum_k a_k^{MR} N_k^{MR} \tag{3}$$

where a_k^{MR} is the molar refractivity group contribution coefficient for the kth molar fragment type and N_k^{MR} is the number of occurrences for the kth molar refractivity fragment type. The AMR descriptor is described in more detail by Viswanadhan and coworkers(Viswanadhan et al. 1989).

The Wang octanol water partition coefficient (XLOGP) is also calculated using a group contribution model:

$$XLOGP = \sum_i a_i A_i + \sum_j b_j B_j \tag{4}$$

where A_i is the occurrence of the ith atom type and B_j is the occurrence of the jth correction factor, a_i is the contribution of the ith atom type and b_j is the contribution of the jth correction factor. The XLOGP descriptor is described in more detail by Wang and coworkers (Wang et al. 2000).

Again XLOGP2 is simply the square of XLOGP:

$$XLOGP2 = XLOGP^2 \tag{5}$$

The hydrophilicity factor is given by

$$Hy = \frac{\left(1 + N_{Hy}\right)\log_2\left(1 + N_{Hy}\right) + N_C\left(\dfrac{1}{A}\log_2\dfrac{1}{A}\right) + \sqrt{\dfrac{N_{Hy}}{A^2}}}{\log_2\left(1 + A\right)} \qquad (6)$$

where N_{Hy} is the number of hydrophilic groups (or the total number of hydrogens attached to oxygen, sulfur, or nitrogen atoms), N_C is the number of carbon atoms, and A is the number of non-hydrogen atoms. The hydrophilicity index is described in more detail on page 225 of the Handbook of Molecular Descriptors (Todeschini and Consonni 2000).

The unsaturation index (UI) is defined as

$$UI = \log_2\left(1 + nDB + nTB + nAB\right) \qquad (7)$$

where nDB = the number of double bonds, nTB = number of triple bonds, and nAB = number of aromatic bonds. The unsaturation index is described in the user manual for Dragon (Talete 2006).

XIV. References

Aguiara, P. F. d., Bourguignon, B., Khotsa, M. S., Massarta, D. L., and Phan-Than-Luub, R. 1995. D-optimal designs. *Chemometrics and Intelligent Laboratory Systems* 30:199-210.

Daylight Chemical Information Systems, Inc. *Simplified Molecular Input Line Entry System*. 2006, http://www.daylight.com/smiles/index.html.

Elsevier MDL. *MDL QSAR Version 2.2*. 2006, http://www.mdl.com/products/predictive/qsar/index.jsp.

Ghose, A. K., Viswanadhan, V. N., and Wendoloski, J. J. 1998. Prediction of Hydrophilic (Lipophilic) Properties of Small Organic Molecules Using Fragmental Methods: An analysis of ALOGP and CLOGP Methods. *J. Phys. Chem.* 102:3762-3772.

Gramatica, P., Corradi, M., and Consonni, V. 2000. Modelling and Prediction of Soil Sorption Coefficients of Non-ionic Organic Pesticides by Molecular Descriptors. *Chemosphere* 41:763-777.

Hall, L. H., and Kier, L. B. 1991. The Molecular Connectivity Chi Indices and Kappa Shape Indices in Structure-Property Relations. In *Reviews of Computational Chemistry*, edited by D. Boyd and K. Lipkowitz. New York: VCH Publishers, Inc., 367-422.

Hall, L. H., and Kier, L. B. 1999. Molecular Connectivity Chi Indices for Database Analysis and Structure-Property Modeling. In *Methods for QSAR Modelling*, edited by J. Devillers.

Kier, L. B. 1987. Inclusion of symmetry as a shape attribute in Kappa index analysis. *Quantit. Struct.-Act. Relat.* 6:8-12.

Kier, L. B., and Hall, L. H. 1976. *Molecular Connectivity in Chemistry and Drug Research*. New York: Academic Press Inc.

Kier, L. B., and Hall, L. H. 1986. *Molecular Connectivity in Structure-Activity Analysis*. New York: John Wiley and Sons.

Kier, L. B., and Hall, L. H. 1999. *Molecule Structure Description: The Electrotopological State*. New York: Academic Press.

Martin, T. M., Harten, P., Venkatapathy, R., Das, S., and Young, D. M. 2008. A Hierarchical Clustering Methodology for the Estimation of Toxicity. *Toxicology Mechanisms and Methods* 18:251–266.

JAMA : A Java Matrix Package. 2005, http://math.nist.gov/javanumerics/jama/.

Talete. *Dragon Version 5.4*. 2006, http://www.talete.mi.it/dragon_net.htm.

Todeschini, R., and Consonni, V. 2000. *Handbook of Molecular Descriptors*. Weinheim, Germany: Wiley-VCH.

Viswanadhan, V. N., Ghose, A. K., Revankar, G. R., and Robins, R. K. 1989. Atomic Physicochemical Parameters for Three Dimensional Structure Directed Quantitative Structure-Activity Relationships. 4. Additional Parameters for Hydrophobic and Dispersive Interactions and Their Application for an Automated Superposition of Certain Naturally Occurring Nucleoside Antibiotics. *J. Chem. Inf. Comput. Sci.* 29:163-172.

Wang, R., Gao, Y., and Lai, L. 2000. Calculating partition coefficient by atom-additive method. *Perspectives in Drug Discovery and Design* 19:47-66.