Complete Properties Codebook for PhysicoChemical Descriptor

❍ **a_acc:** Number of hydrogen bond acceptor atoms (not counting acidic atoms but counting atoms that are both hydrogen bond donors and acceptors such as -OH).

❍ **a_acid:** Number of acidic atoms.

❍ **a_aro:** Number of aromatic atoms.

❍ **a_base:** Number of basic atoms.

❍ **a_count:** Number of atoms (including implicit hydrogens). This is calculated as the sum of $(1 + h_i)$ over all non-trivial atoms i.

❍ **a_don:** Number of hydrogen bond donor atoms (not counting basic atoms but counting atoms that are both hydrogen bond donors and acceptors such as -OH).

❍ **a_heavy:** Number of heavy atoms #$\{Z_i \mid Z_i > 1\}$.

❍ **a_hyd:** Number of hydrophobic atoms.

❍ **a_IC:** Atom information content (total). This is calculated to be a_ICM times n.

❍ **a_ICM:** Atom information content (mean). This is the entropy of the element distribution in the molecule (including implicit hydrogens but not lone pair pseudo-atoms). Let n_i be the number of occurrences of atomic number i in the molecule. Let $p_i = n_i / n$ where n is the sum of the n_i. The value of a_ICM is the negative of the sum over all i of $p_i \log p_i$.

❍ **a_nB:** Number of boron atoms: #$\{Z_i \mid Z_i = 5\}$.

❍ **a_nBr:** Number of bromine atoms: #$\{Z_i \mid Z_i = 35\}$.

○ **a_nC:** Number of carbon atoms: #{Zi | Zi = 6}.

○ **a_nCl:** Number of chlorine atoms: #{Zi | Zi = 17}.

○ **a_nF:** Number of fluorine atoms: #{Zi | Zi = 9}.

○ **a_nH:** Number of hydrogen atoms (including implicit hydrogens). This is calculated as the sum of hi over all non-trivial atoms i plus the number of non-trivial hydrogen atoms.

○ **a_nI:** Number of iodine atoms: #{Zi | Zi = 53}.

○ **a_nN:** Number of nitrogen atoms: #{Zi | Zi = 7}.

○ **a_nO:** Number of oxygen atoms: #{Zi | Zi = 8}.

○ **a_nP:** Number of phosphorus atoms: #{Zi | Zi = 15}.

○ **a_nS:** Number of sulfur atoms: #{Zi | Zi = 16}.

○ **AM1_dipole:** The dipole moment calculated using the AM1 Hamiltonian. [*Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.*]

○ **AM1_E:** The total SCF energy (kcal/mol) calculated using the AM1 Hamiltonian. [*Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.*]

○ **AM1_Eele:** The electronic energy (kcal/mol) calculated using the AM1 Hamiltonian. [*Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.*]

○ **AM1_HF:** The heat of formation (kcal/mol) calculated using the AM1 Hamiltonian. [*Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.*]

○ **AM1_HOMO:** The energy (eV) of the Highest Occupied Molecular Orbital calculated using the AM1 Hamiltonian. [*Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.*]

○ **AM1_IP:** The ionization potential (kcal/mol) calculated using the AM1 Hamiltonian. [*Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.*]

○ **AM1_LUMO:** The energy (eV) of the Lowest Unoccupied Molecular Orbital calculated using the AM1 Hamiltonian. [*Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.*]

○ **apol:** Sum of the atomic polarizabilities (including implicit hydrogens) with polarizabilities taken from CRC Handbook of Chemistry and Physics. CRC Press (1994).

○ **ASA:** Water accessible surface area calculated using a radius of 1.4 A for the water molecule. A polyhedral representation is used for each atom in calculating the surface area.

○ **ASA+:** Water accessible surface area of all atoms with positive partial charge (strictly greater than 0).

○ **ASA-:** Water accessible surface area of all atoms with negative partial charge (strictly less than 0).

○ **ASA_H:** Water accessible surface area of all hydrophobic ($|qi|<0.2$) atoms.

○ **ASA_P:** Water accessible surface area of all polar ($|qi|>=0.2$) atoms.

○ **b_1rotN:** Number of rotatable single bonds. Conjugated single bonds are not included (e.g. ester and peptide bonds).

○ **b_1rotR:** Fraction of rotatable single bonds: b_1rotN divided by b_heavy.

○ **b_ar:** Number of aromatic bonds.

○ **b_count:** Number of bonds (including implicit hydrogens). This is calculated as the sum of (di/2 + hi) over all non-trivial atoms i.

○ **b_double:** Number of double bonds. Aromatic bonds are not considered to be

double bonds.

❍ **b_heavy:** Number of bonds between heavy atoms.

❍ **b_rotN:** Number of rotatable bonds. A bond is rotatable if it has order 1, is not in a ring, and has at least two heavy neighbors.

❍ **b_rotR:** Fraction of rotatable bonds: b_rotN divided by b_heavy.

❍ **b_single:** Number of single bonds (including implicit hydrogens). Aromatic bonds are not considered to be single bonds.

❍ **b_triple:** Number of triple bonds. Aromatic bonds are not considered to be triple bonds.

❍ **balabanJ:** Balaban's connectivity topological index. [*Balaban, A.T.; Highly Discriminating Distance-Based Topological Index; Chemical Physics Letters 89 No. 5 (1982) 399-404.*]

❍ **BCUT_PEOE_0:** The BCUT descriptors [*Pearlman, R.S., Smith, K.M.; Novel Software Tools for Chemical Diversity; Persp. Drug. Disc. Des. 9/10/11 (1998) 339-353.*] are calculated from the eigenvalues of a modified adjacency matrix. Each ij entry of the adjacency matrix takes the value 1/sqrt(bij) where bij is the formal bond order between bonded atoms i and j. The diagonal takes the value of the PEOE partial charges. The resulting eigenvalues are sorted and the smallest, 1/3-ile, 2/3-ile and largest eigenvalues are reported.

❍ **BCUT_PEOE_1:** The BCUT descriptors [*Pearlman, R.S., Smith, K.M.; Novel Software Tools for Chemical Diversity; Persp. Drug. Disc. Des. 9/10/11 (1998) 339-353.*] are calculated from the eigenvalues of a modified adjacency matrix. Each ij entry of the adjacency matrix takes the value 1/sqrt(bij) where bij is the formal bond order between bonded atoms i and j. The diagonal takes the value of the PEOE partial charges. The resulting eigenvalues are sorted and the smallest, 1/3-ile, 2/3-ile and largest eigenvalues are reported.

❍ **BCUT_PEOE_2:** The BCUT descriptors [*Pearlman, R.S., Smith, K.M.; Novel Software Tools for Chemical Diversity; Persp. Drug. Disc. Des. 9/10/11 (1998) 339-353.*] are calculated from the eigenvalues of a modified adjacency matrix. Each ij entry of the adjacency matrix takes the value 1/sqrt(bij) where bij is the formal bond order between bonded atoms i and j. The diagonal takes the value of the PEOE partial charges. The resulting eigenvalues are sorted and the smallest, 1/3-ile, 2/3-ile and largest eigenvalues are reported.

❍ **BCUT_PEOE_3:** The BCUT descriptors [*Pearlman, R.S., Smith, K.M.; Novel Software Tools for Chemical Diversity; Persp. Drug. Disc. Des. 9/10/11 (1998) 339-353.*] are calculated from the eigenvalues of a modified adjacency matrix. Each ij entry of the adjacency matrix takes the value 1/sqrt(bij) where bij is the formal bond order between bonded atoms i and j. The diagonal takes the value of the PEOE partial charges. The resulting eigenvalues are sorted and the smallest, 1/3-ile, 2/3-ile and largest eigenvalues are reported.

❍ **BCUT_SLOGP_0:** The BCUT descriptors using atomic contribution to logP (using the Wildman and Crippen SlogP method) instead of partial charge.

❍ **BCUT_SLOGP_1:** The BCUT descriptors using atomic contribution to logP (using the Wildman and Crippen SlogP method) instead of partial charge.

❍ **BCUT_SLOGP_2:** The BCUT descriptors using atomic contribution to logP (using the Wildman and Crippen SlogP method) instead of partial charge.

❍ **BCUT_SLOGP_3:** The BCUT descriptors using atomic contribution to logP (using the Wildman and Crippen SlogP method) instead of partial charge.

❍ **BCUT_SMR_0:** The BCUT descriptors using atomic contribution to molar refractivity (using the Wildman and Crippen SMR method) instead of partial charge.

❍ **BCUT_SMR_1:** The BCUT descriptors using atomic contribution to molar refractivity (using the Wildman and Crippen SMR method) instead of partial charge.

○ **BCUT_SMR_2:** The BCUT descriptors using atomic contribution to molar refractivity (using the Wildman and Crippen SMR method) instead of partial charge.

○ **BCUT_SMR_3:** The BCUT descriptors using atomic contribution to molar refractivity (using the Wildman and Crippen SMR method) instead of partial charge.

○ **bpol:** Sum of the absolute value of the difference between atomic polarizabilities of all bonded atoms in the molecule (including implicit hydrogens) with polarizabilities taken from CRC Handbook of Chemistry and Physics. CRC Press (1994).

○ **CASA+:** Positive charge weighted surface area, ASA+ times max { q_i > 0 }. **[**_Stanton, D., Jurs, P.; Development and Use of Charged Partial Surface-Area Structural Descriptors in Computer-Assisted Quantitative Structure-Property Relationship Studies; Anal. Chem. 62 (1990) 2323-2329._**]**

○ **CASA-:** Negative charge weighted surface area, ASA- times max { q_i < 0 }. **[**_Stanton, D., Jurs, P.; Development and Use of Charged Partial Surface-Area Structural Descriptors in Computer-Assisted Quantitative Structure-Property Relationship Studies; Anal. Chem. 62 (1990) 2323-2329._**]**

○ **chi0:** Atomic connectivity index (order 0) from **[**_Hall, L.H., Kier, L.B.; The Molecular Connectivity Chi Indices and Kappa Shape Indices in Structure-Property Modeling; Reviews of Computational Chemistry 2 (1991)._**]** and **[**_Hall, L.H., Kier, L.B.; The Nature of Structure-Activity Relationships and Their Relation to Molecular Connectivity; Eur. J. Med. Chem 12 (1977) 307._**]**. This is calculated as the sum of 1/sqrt(di) over all heavy atoms i with di > 0.

○ **chi0_C:** Carbon connectivity index (order 0). This is calculated as the sum of 1/sqrt(di) over all carbon atoms i with di > 0.

○ **chi0v:** Atomic valence connectivity index (order 0) from **[**_Hall, L.H., Kier, L.B.; The Molecular Connectivity Chi Indices and Kappa Shape Indices in_

Structure-Property Modeling; Reviews of Computational Chemistry 2 (1991).] and [*Hall, L.H., Kier, L.B.; The Nature of Structure-Activity Relationships and Their Relation to Molecular Connectivity; Eur. J. Med. Chem 12 (1977) 307.*]. This is calculated as the sum of 1/sqrt(vi) over all heavy atoms i with vi > 0.

❍ **chi0v_C:** Carbon valence connectivity index (order 0). This is calculated as the sum of 1/sqrt(vi) over all carbon atoms i with vi > 0.

❍ **chi1:** Atomic connectivity index (order 1) from [*Hall, L.H., Kier, L.B.; The Molecular Connectivity Chi Indices and Kappa Shape Indices in Structure-Property Modeling; Reviews of Computational Chemistry 2 (1991).*] and [*Hall, L.H., Kier, L.B.; The Nature of Structure-Activity Relationships and Their Relation to Molecular Connectivity; Eur. J. Med. Chem 12 (1977) 307.*]. This is calculated as the sum of 1/sqrt(didj) over all bonds between heavy atoms i and j where i < j.

❍ **chi1_C:** Carbon connectivity index (order 1). This is calculated as the sum of 1/sqrt(didj) over all bonds between carbon atoms i and j where i < j.

❍ **chi1v:** Atomic valence connectivity index (order 1) from [*Hall, L.H., Kier, L.B.; The Molecular Connectivity Chi Indices and Kappa Shape Indices in Structure-Property Modeling; Reviews of Computational Chemistry 2 (1991).*] and [*Hall, L.H., Kier, L.B.; The Nature of Structure-Activity Relationships and Their Relation to Molecular Connectivity; Eur. J. Med. Chem 12 (1977) 307.*]. This is calculated as the sum of 1/sqrt(vivj) over all bonds between heavy atoms i and j where i < j.

❍ **chi1v_C:** Carbon valence connectivity index (order 1). This is calculated as the sum of 1/sqrt(vivj) over all bonds between carbon atoms i and j where i < j.

❍ **chiral:** The number of chiral centers.

❍ **chiral_u:** The number of unconstrained chiral centers.

❍ **DASA:** Absolute value of the difference between ASA+ and ASA-.

❍ **DCASA:** Absolute value of the difference between CASA+ and CASA-. [*Stanton, D., Jurs, P.; Development and Use of Charged Partial Surface-Area Structural Descriptors in Computer-Assisted Quantitative Structure-Property Relationship Studies; Anal. Chem. 62 (1990) 2323-2329.*]

❍ **dens:** Mass density: molecular weight divided by van der Waals volume as calculated in the vol descriptor.

❍ **density:** Molecular mass density: Weight divided by vdw_vol (amu/A^3).

❍ **diameter:** Largest value in the distance matrix. [*Petitjean, M.; Applications of the Radius-Diameter Diagram to the Classification of Topological and Geometrical Shapes of Chemical Compounds; J. Chem. Inf. Comput. Sci. 32 (1992) 331-337.*]

❍ **dipole:** Dipole moment calculated from the partial charges of the molecule.

❍ **dipoleX:** The x component of the dipole moment (external coordinates).

❍ **dipoleY:** The y component of the dipole moment (external coordinates).

❍ **dipoleZ:** The z component of the dipole moment (external coordinates).

❍ **E:** Value of the potential energy.

❍ **E_ang:** Angle bend potential energy.

❍ **E_ele:** Electrostatic component of the potential energy.

❍ **E_nb:** Value of the potential energy with all bonded terms disabled.

❍ **E_oop:** Out-of-plane potential energy.

❍ **E_rele:** Electrostatic interaction energy (external reference frame: x3d) between the stored molecule and the atoms currently loaded.

❍ **E_rnb:** Non-bonded interaction energy (external reference frame: x3d) between the stored molecule and a "receptor" currently loaded.

❍ **E_rsol:** Solvation free energy difference (external reference frame: x3d). Let L be the free energy of solvation of the stored molecule (ligand), R be the free energy of solvation of the atoms currently loaded (receptor), and G be the free energy of solvation of the RL complex. Consequently, the returned value is G - L - R.

❍ **E_rvdw:** van der Waals interaction energy (external reference frame: x3d) between the stored molecule and the atoms currently loaded.

❍ **E_sol:** Solvation energy.

❍ **E_stb:** Bond stretch-bend cross-term potential energy.

❍ **E_str:** Bond stretch potential energy.

❍ **E_strain:** Local strain energy: the current energy minus the value of the energy at a near local minimum. The current energy is calculated as for the E descriptor. The local minimum energy is the value of the E descriptor after first performing an energy minimization.

❍ **E_tor:** Torsion (proper and improper) potential energy.

❍ **E_vdw:** van der Waals component of the potential energy.

❍ **FASA+:** Fractional ASA+ calculated as ASA+ / ASA.

❍ **FASA-:** Fractional ASA- calculated as ASA- / ASA.

❍ **FASA_H:** Fractional ASA_H calculated as ASA_H / ASA.

❍ **FASA_P:** Fractional ASA_P calculated as ASA_P / ASA.

○ **FCASA+:** Fractional CASA+ calculated as CASA+ / ASA.

○ **FCASA-:** Fractional CASA- calculated as CASA- / ASA.

○ **FCharge:** Total charge of the molecule (sum of formal charges).

○ **GCUT_PEOE_0:** The GCUT descriptors are calculated from the eigenvalues of a modified graph distance adjacency matrix. Each ij entry of the adjacency matrix takes the value 1/sqr(dij) where dij is the (modified) graph distance between atoms i and j. The diagonal takes the value of the PEOE partial charges. The resulting eigenvalues are sorted and the smallest, 1/3-ile, 2/3-ile and largest eigenvalues are reported.

○ **GCUT_PEOE_1:** The GCUT descriptors are calculated from the eigenvalues of a modified graph distance adjacency matrix. Each ij entry of the adjacency matrix takes the value 1/sqr(dij) where dij is the (modified) graph distance between atoms i and j. The diagonal takes the value of the PEOE partial charges. The resulting eigenvalues are sorted and the smallest, 1/3-ile, 2/3-ile and largest eigenvalues are reported.

○ **GCUT_PEOE_2:** The GCUT descriptors are calculated from the eigenvalues of a modified graph distance adjacency matrix. Each ij entry of the adjacency matrix takes the value 1/sqr(dij) where dij is the (modified) graph distance between atoms i and j. The diagonal takes the value of the PEOE partial charges. The resulting eigenvalues are sorted and the smallest, 1/3-ile, 2/3-ile and largest eigenvalues are reported.

○ **GCUT_PEOE_3:** The GCUT descriptors are calculated from the eigenvalues of a modified graph distance adjacency matrix. Each ij entry of the adjacency matrix takes the value 1/sqr(dij) where dij is the (modified) graph distance between atoms i and j. The diagonal takes the value of the PEOE partial charges. The resulting eigenvalues are sorted and the smallest, 1/3-ile, 2/3-ile and largest eigenvalues are reported.

○ **GCUT_SLOGP_0:** The GCUT descriptors using atomic contribution to logP (using the Wildman and Crippen SlogP method) instead of partial charge.

○ **GCUT_SLOGP_1:** The GCUT descriptors using atomic contribution to logP (using the Wildman and Crippen SlogP method) instead of partial charge.

○ **GCUT_SLOGP_2:** The GCUT descriptors using atomic contribution to logP (using the Wildman and Crippen SlogP method) instead of partial charge.

○ **GCUT_SLOGP_3:** The GCUT descriptors using atomic contribution to logP (using the Wildman and Crippen SlogP method) instead of partial charge.

○ **GCUT_SMR_0:** The GCUT descriptors using atomic contribution to molar refractivity (using the Wildman and Crippen SMR method) instead of partial charge.

○ **GCUT_SMR_1:** The GCUT descriptors using atomic contribution to molar refractivity (using the Wildman and Crippen SMR method) instead of partial charge.

○ **GCUT_SMR_2:** The GCUT descriptors using atomic contribution to molar refractivity (using the Wildman and Crippen SMR method) instead of partial charge.

○ **GCUT_SMR_3:** The GCUT descriptors using atomic contribution to molar refractivity (using the Wildman and Crippen SMR method) instead of partial charge.

○ **glob:** Globularity, or inverse condition number (smallest eigenvalue divided by the largest eigenvalue) of the covariance matrix of atomic coordinates. A value of 1 indicates a perfect sphere while a value of 0 indicates a two- or one-dimensional object.

○ **Hydro_IMGT:** Hydropathy value of amino acids, as reported by IMGT itself.

○ **Kier1:** First kappa shape index: $(n-1)^2 / m^2$. [*Hall, L.H., Kier, L.B.; The Molecular Connectivity Chi Indices and Kappa Shape Indices in Structure-Property Modeling; Reviews of Computational Chemistry 2 (1991).*]

○ **Kier2:** Second kappa shape index: $(n-1)^2 / m^2$. [*Hall, L.H., Kier, L.B.; The*

Molecular Connectivity Chi Indices and Kappa Shape Indices in Structure-Property Modeling; Reviews of Computational Chemistry 2 (1991).]

❍ **Kier3:** Third kappa shape index: (n-1) (n-3)^2 / p3^2 for odd n, and (n-3) (n-2)^2 / p3^2 for even n. [*Hall, L.H., Kier, L.B.; The Molecular Connectivity Chi Indices and Kappa Shape Indices in Structure-Property Modeling; Reviews of Computational Chemistry 2 (1991).*]

❍ **KierA1:** First alpha modified shape index: s (s-1)^2 / m^2 where s = n + a. [*Hall, L.H., Kier, L.B.; The Molecular Connectivity Chi Indices and Kappa Shape Indices in Structure-Property Modeling; Reviews of Computational Chemistry 2 (1991).*]

❍ **KierA2:** Second alpha modified shape index: s (s-1)^2 / m^2 where s = n + a. [*Hall, L.H., Kier, L.B.; The Molecular Connectivity Chi Indices and Kappa Shape Indices in Structure-Property Modeling; Reviews of Computational Chemistry 2 (1991).*]

❍ **KierA3:** Third alpha modified shape index: (s-1) (s-3)^2 / p3^2 for odd n, and (s-3) (s-2)^2 / p3^2 for even n where s = n + a. [*Hall, L.H., Kier, L.B.; The Molecular Connectivity Chi Indices and Kappa Shape Indices in Structure-Property Modeling; Reviews of Computational Chemistry 2 (1991).*]

❍ **KierFlex:** Kier molecular flexibility index: (KierA1) (KierA2) / n. [*Hall, L.H., Kier, L.B.; The Molecular Connectivity Chi Indices and Kappa Shape Indices in Structure-Property Modeling; Reviews of Computational Chemistry 2 (1991).*]

❍ **lip_acc:** The number of O and N atoms.

❍ **lip_don:** The number of OH and NH atoms.

❍ **lip_druglike:** One if and only if lip_violation < 2 otherwise zero.

❍ **lip_violation:** The number of violations of Lipinski's Rule of Five. [*Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J.; Experimental and Computational Approaches to Estimate Solubility and*

Permeability in Drug Discovery and Development Settings; Adv. Drug Deliv. Rev. 23 1997) 3-25.]

❍ **logP(o/w):** Log of the octanol/water partition coefficient (including implicit hydrogens). This property is calculated from a linear atom type model[*Labute, P.; MOE LogP(Octanol/Water) Model unpublished. Source code in $MOE/lib/svl/quasar.svl/q_logp.svl (1998).*] with r^2 = 0.931, RMSE=0.393 on 1,827 molecules.

❍ **logS:** Log of the aqueous solubility (mol/L). This property is calculated from an atom contribution linear atom type model [*Hou, T.J., Xia, K., Zhang, W., Xu, X.J.; ADME Evaluation in Drug Discovery. 4. Prediction of Aqueous Solubility Based on Atom Contribution Approach; J. Chem. Inf. Comput. Sci. 44(2004) 266-275.*] with r^2 = 0.90, ~1,200 molecules.

❍ **MNDO_dipole:** The dipole moment calculated using the MNDO Hamiltonian. [*Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.*]

❍ **MNDO_E:** The total SCF energy (kcal/mol) calculated using the MNDO Hamiltonian. [*Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.*]

❍ **MNDO_Eele:** The electronic energy (kcal/mol) calculated using the MNDO Hamiltonian. [*Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.*]

❍ **MNDO_HF:** The heat of formation (kcal/mol) calculated using the MNDO Hamiltonian. [*Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.*]

❍ **MNDO_HOMO:** The energy (eV) of the Highest Occupied Molecular Orbital calculated using the MNDO Hamiltonian. [*Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.*]

❍ **MNDO_IP:** The ionization potential (kcal/mol) calculated using the MNDO Hamiltonian. [*Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.*]

❍ **MNDO_LUMO:** The energy (eV) of the Lowest Unoccupied Molecular Orbital calculated using the MNDO Hamiltonian. [*Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.*]

○ **mr:** Molecular refractivity (including implicit hydrogens). This property is calculated from an 11 descriptor linear model [*Labute, P.; MOE Molar Refractivity Model unpublished. Source code in $MOE/lib/svl/quasar.svl/q_mref.svl (1998).*] with r2 = 0.997, RMSE = 0.168 on 1,947 small molecules.

○ **mutagenic:** Indicator of the presence of potentially toxic groups. A non-zero value indicates that the molecule contains a mutagenic group. The table of mutagenic groups is based on the Kazius set [*Kazius, J., McGuire, R., Bursi, R.; Derivation and Validation of Toxicophores for Mutagenicity Prediction; J. Med. Chem. 48 (2005) 312-320.*].

○ **nmol:** The number of molecules (connected components).

○ **npr1:** Normalized PMI ratio pmi1/pmi3.

○ **npr2:** Normalized PMI ratio pmi2/pmi3.

○ **opr_brigid:** The number of rigid bonds from Oprea, Tudor I.; Property Distribution of Drug-Related Chemical Databases; J. Comp. Aid. Mol. Des. 14 (2000) 251-264.

○ **opr_leadlike:** One if and only if opr_violation < 2 otherwise zero.

○ **opr_nring:** The number of ring bonds from Oprea, Tudor I.; Property Distribution of Drug-Related Chemical Databases; J. Comp. Aid. Mol. Des. 14 (2000) 251-264.

○ **opr_nrot:** The number of rotatable bonds from Oprea, Tudor I.; Property Distribution of Drug-Related Chemical Databases; J. Comp. Aid. Mol. Des. 14 (2000) 251-264.

○ **opr_violation:** The number of violations of Oprea's lead-like test. [*Oprea, Tudor I.; Property Distribution of Drug-Related Chemical Databases; J. Comp. Aid. Mol. Des. 14 (2000) 251-264.*]

❍ **PC+:** Total positive partial charge: the sum of the positive q_i.

❍ **PC-:** Total negative partial charge: the sum of the negative q_i.

❍ **PEOE_PC+:** Total positive partial charge: the sum of the positive q_i.

❍ **PEOE_PC-:** Total negative partial charge: the sum of the negative q_i.

❍ **PEOE_RPC+:** Relative positive partial charge: the largest positive q_i divided by the sum of the positive q_i.

❍ **PEOE_RPC-:** Relative negative partial charge: the smallest negative q_i divided by the sum of the negative q_i.

❍ **PEOE_VSA+0:** Sum of v_i where q_i is in the range $[0.00, 0.05)$.

❍ **PEOE_VSA+1:** Sum of v_i where q_i is in the range $[0.05, 0.10)$.

❍ **PEOE_VSA+2:** Sum of v_i where q_i is in the range $[0.10, 0.15)$.

❍ **PEOE_VSA+3:** Sum of v_i where q_i is in the range $[0.15, 0.20)$.

❍ **PEOE_VSA+4:** Sum of v_i where q_i is in the range $[0.20, 0.25)$.

❍ **PEOE_VSA+5:** Sum of v_i where q_i is in the range $[0.25, 0.30)$.

❍ **PEOE_VSA+6:** Sum of v_i where q_i is greater than 0.3.

❍ **PEOE_VSA-0:** Sum of v_i where q_i is in the range $[-0.05, 0.00)$.

❍ **PEOE_VSA-1:** Sum of v_i where q_i is in the range $[-0.10, -0.05)$.

❍ **PEOE_VSA-2:** Sum of v_i where q_i is in the range $[-0.15, -0.10)$.

❍ **PEOE_VSA-3:** Sum of v_i where q_i is in the range $[-0.20, -0.15)$

○ **PEOE_VSA-4:** Sum of vi where qi is in the range [-0.25,-0.20).

○ **PEOE_VSA-5:** Sum of vi where qi is in the range [-0.30,-0.25).

○ **PEOE_VSA-6:** Sum of vi where qi is less than -0.30.

○ **PEOE_VSA_FHYD:** Fractional hydrophobic van der Waals surface area. This is the sum of the vi such that |qi| is less than or equal to 0.2 divided by the total surface area. The vi are calculated using a connection table approximation.

○ **PEOE_VSA_FNEG:** Fractional negative van der Waals surface area. This is the sum of the vi such that qi is negative divided by the total surface area. The vi are calculated using a connection table approximation.

○ **PEOE_VSA_FPNEG:** Fractional negative polar van der Waals surface area. This is the sum of the vi such that qi is less than -0.2 divided by the total surface area. The vi are calculated using a connection table approximation.

○ **PEOE_VSA_FPOL:** Fractional polar van der Waals surface area. This is the sum of the vi such that |qi| is greater than 0.2 divided by the total surface area. The vi are calculated using a connection table approximation.

○ **PEOE_VSA_FPOS:** Fractional positive van der Waals surface area. This is the sum of the vi such that qi is non-negative divided by the total surface area. The vi are calculated using a connection table approximation.

○ **PEOE_VSA_FPPOS:** Fractional positive polar van der Waals surface area. This is the sum of the vi such that qi is greater than 0.2 divided by the total surface area. The vi are calculated using a connection table approximation.

○ **PEOE_VSA_HYD:** Total hydrophobic van der Waals surface area. This is the sum of the vi such that |qi| is less than or equal to 0.2. The vi are

calculated using a connection table approximation.

○ **PEOE_VSA_NEG:** Total negative van der Waals surface area. This is the sum of the vi such that qi is negative. The vi are calculated using a connection table approximation.

○ **PEOE_VSA_PNEG:** Total negative polar van der Waals surface area. This is the sum of the vi such that qi is less than -0.2. The vi are calculated using a connection table approximation.

○ **PEOE_VSA_POL:** Total polar van der Waals surface area. This is the sum of the vi such that |qi| is greater than 0.2. The vi are calculated using a connection table approximation.

○ **PEOE_VSA_POS:** Total positive van der Waals surface area. This is the sum of the vi such that qi is non-negative. The vi are calculated using a connection table approximation.

○ **PEOE_VSA_PPOS:** Total positive polar van der Waals surface area. This is the sum of the vi such that qi is greater than 0.2. The vi are calculated using a connection table approximation.

○ **petitjean:** Value of (diameter - radius) / diameter.

○ **petitjeanSC:** Petitjean graph Shape Coefficient as defined in [*Petitjean, M.; Applications of the Radius-Diameter Diagram to the Classification of Topological and Geometrical Shapes of Chemical Compounds; J. Chem. Inf. Comput. Sci. 32 (1992) 331-337.*]: (diameter - radius) / radius.

○ **PM3_dipole:** The dipole moment calculated using the PM3 Hamiltonian. [*Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.*]

○ **PM3_E:** The total SCF energy (kcal/mol) calculated using the PM3 Hamiltonian. [*Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.*]

○ **PM3_Eele:** The electronic energy (kcal/mol) calculated using the PM3 Hamiltonian. [*Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.*]

❍ **PM3_HF:** The heat of formation (kcal/mol) calculated using the PM3 Hamiltonian. [*Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.*]

❍ **PM3_HOMO:** The energy (eV) of the Highest Occupied Molecular Orbital calculated using the PM3 Hamiltonian. [*Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.*]

❍ **PM3_IP:** The ionization potential (kcal/mol) calculated using the PM3 Hamiltonian. [*Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.*]

❍ **PM3_LUMO:** The energy (eV) of the Lowest Unoccupied Molecular Orbital calculated using the PM3 Hamiltonian. [*Stewart, J.J.P.; MOPAC Manual (Seventh Edition); 1993.*]

❍ **pmi:** Principal moment of inertia.

❍ **pmi1:** First diagonal element of diagonalized moment of inertia tensor.

❍ **pmi2:** Second diagonal element of diagonalized moment of inertia tensor.

❍ **pmi3:** Third diagonal element of diagonalized moment of inertia tensor.

❍ **pmiX:** x component of the principal moment of inertia (external coordinates).

❍ **pmiY:** y component of the principal moment of inertia (external coordinates).

❍ **pmiZ:** z component of the principal moment of inertia (external coordinates).

❍ **Q_PC+:** Total positive partial charge: the sum of the positive q_i.

❍ **Q_PC-:** Total negative partial charge: the sum of the negative q_i.

- **Q_RPC+:** Relative positive partial charge: the largest positive q_i divided by the sum of the positive q_i.

- **Q_RPC-:** Relative negative partial charge: the smallest negative q_i divided by the sum of the negative q_i.

- **Q_VSA_FHYD:** Fractional hydrophobic van der Waals surface area. This is the sum of the v_i such that $|q_i|$ is less than or equal to 0.2 divided by the total surface area. The v_i are calculated using a connection table approximation.

- **Q_VSA_FNEG:** Fractional negative van der Waals surface area. This is the sum of the v_i such that q_i is negative divided by the total surface area. The v_i are calculated using a connection table approximation.

- **Q_VSA_FPNEG:** Fractional negative polar van der Waals surface area. This is the sum of the v_i such that q_i is less than -0.2 divided by the total surface area. The v_i are calculated using a connection table approximation.

- **Q_VSA_FPOL:** Fractional polar van der Waals surface area. This is the sum of the v_i such that $|q_i|$ is greater than 0.2 divided by the total surface area. The v_i are calculated using a connection table approximation.

- **Q_VSA_FPOS:** Fractional positive van der Waals surface area. This is the sum of the v_i such that q_i is non-negative divided by the total surface area. The v_i are calculated using a connection table approximation.

- **Q_VSA_FPPOS:** Fractional positive polar van der Waals surface area. This is the sum of the v_i such that q_i is greater than 0.2 divided by the total surface area. The v_i are calculated using a connection table approximation.

- **Q_VSA_HYD:** Total hydrophobic van der Waals surface area. This is the sum of the v_i such that $|q_i|$ is less than or equal to 0.2. The v_i are calculated using a connection table approximation.

❍ **Q_VSA_NEG:** Total negative van der Waals surface area. This is the sum of the vi such that qi is negative. The vi are calculated using a connection table approximation.

❍ **Q_VSA_PNEG:** Total negative polar van der Waals surface area. This is the sum of the vi such that qi is less than -0.2. The vi are calculated using a connection table approximation.

❍ **Q_VSA_POL:** Total polar van der Waals surface area. This is the sum of the vi such that |qi| is greater than 0.2. The vi are calculated using a connection table approximation.

❍ **Q_VSA_POS:** Total positive van der Waals surface area. This is the sum of the vi such that qi is non-negative. The vi are calculated using a connection table approximation.

❍ **Q_VSA_PPOS:** Total positive polar van der Waals surface area. This is the sum of the vi such that qi is greater than 0.2. The vi are calculated using a connection table approximation.

❍ **radius:** Radius of gyration.

❍ **reactive:** Indicator of the presence of reactive groups. A non-zero value indicates that the molecule contains a reactive group. The table of reactive groups is based on the Oprea set [*Oprea, Tudor I.; Property Distribution of Drug-Related Chemical Databases; J. Comp. Aid. Mol. Des. 14 (2000) 251-264.*]and includes metals, phospho-, N/O/S-N/O/S single bonds, thiols, acyl halides, Michael Acceptors, azides, esters, etc.

❍ **rgyr:** Radius of gyration.

❍ **rings:** The number of rings.

❍ **RPC+:** Relative positive partial charge: the largest positive qi divided by the sum of the positive qi.

❍ **RPC-:** Relative negative partial charge: the smallest negative qi divided

by the sum of the negative qi.

❍ **rsynth:** A value in **[*0,1*]** indicating the synthetic reasonableness, or feasibility, of the chemical structure. A value of 0 means it is unlikely that the molecule can be synthesized while a value of 1 means that it is likely that the molecule can be synthesized. The value reflects the fraction of heavy atoms in the molecule that can be traced back to starting materials fragments resulting from retrosynthetic disconnection rules.

❍ **SlogP:** Log of the octanol/water partition coefficient (including implicit hydrogens). This property is an atomic contribution model **[*Wildman, S.A., Crippen, G.M.; Prediction of Physicochemical Parameters by Atomic Contributions; J. Chem. Inf. Comput. Sci. 39 No. 5 (1999) 868-873.*]** that calculates logP from the given structure; i.e. the correct protonation state (washed structures). Results may vary from the logP(o/w) descriptor. The training set for SlogP was ~7000 structures.

❍ **SlogP_VSA0:** Sum of vi such that Li <= -0.4.

❍ **SlogP_VSA1:** Sum of vi such that Li is in (-0.4,-0.2].

❍ **SlogP_VSA2:** Sum of vi such that Li is in (-0.2,0].

❍ **SlogP_VSA3:** Sum of vi such that Li is in (0,0.1].

❍ **SlogP_VSA4:** Sum of vi such that Li is in (0.1,0.15].

❍ **SlogP_VSA5:** Sum of vi such that Li is in (0.15,0.20].

❍ **SlogP_VSA6:** Sum of vi such that Li is in (0.20,0.25].

❍ **SlogP_VSA7:** Sum of vi such that Li is in (0.25,0.30].

❍ **SlogP_VSA8:** Sum of vi such that Li is in (0.30,0.40].

○ **SlogP_VSA9:** Sum of vi such that Li > 0.40.

○ **SMR:** Molecular refractivity (including implicit hydrogens). This property is an atomic contribution model [*Wildman, S.A., Crippen, G.M.; Prediction of Physicochemical Parameters by Atomic Contributions; J. Chem. Inf. Comput. Sci. 39 No. 5 (1999) 868-873.*] that assumes the correct protonation state (washed structures). The model was trained on ~7000 structures and results may vary from the mr descriptor.

○ **SMR_VSA0:** Sum of vi such that Ri is in **[*0,0.11*]**.

○ **SMR_VSA1:** Sum of vi such that Ri is in (0.11,0.26].

○ **SMR_VSA2:** Sum of vi such that Ri is in (0.26,0.35].

○ **SMR_VSA3:** Sum of vi such that Ri is in (0.35,0.39].

○ **SMR_VSA4:** Sum of vi such that Ri is in (0.39,0.44].

○ **SMR_VSA5:** Sum of vi such that Ri is in (0.44,0.485].

○ **SMR_VSA6:** Sum of vi such that Ri is in (0.485,0.56].

○ **SMR_VSA7:** Sum of vi such that Ri > 0.56.

○ **std_dim1:** Standard dimension 1: the square root of the largest eigenvalue of the covariance matrix of the atomic coordinates. A standard dimension is equivalent to the standard deviation along a principal component axis.

○ **std_dim2:** Standard dimension 2: the square root of the second largest eigenvalue of the covariance matrix of the atomic coordinates. A standard dimension is equivalent to the standard deviation along a principal component axis.

○ **std_dim3:** Standard dimension 3: the square root of the third largest eigenvalue of the covariance matrix of the atomic coordinates. A standard

dimension is equivalent to the standard deviation along a principal component axis.

❍ **TPSA:** Polar surface area (A^2) calculated using group contributions to approximate the polar surface area from connection table information only. The parameterization is that of Ertl et al. [*Ertl, P., Rohde, B., Selzer, P.; Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties; J. Med. Chem. 43 (2000) 3714-3717.*]

❍ **VAdjEq:** Vertex adjacency information (equality): $-(1-f)\log_2(1-f) - f\log_2(f)$ where $f = (n^2 - m) / n^2$, n is the number of heavy atoms and m is the number of heavy-heavy bonds. If f is not in the open interval (0,1), then 0 is returned.

❍ **VAdjMa:** Vertex adjacency information (magnitude): $1 + \log_2(m)$ where m is the number of heavy-heavy bonds. If m is zero, then zero is returned.

❍ **VDistEq:** If m is the sum of the distance matrix entries then VdistEq is defined to be the sum of log2 m - pi log2 pi / m where pi is the number of distance matrix entries equal to i.

❍ **VDistMa:** If m is the sum of the distance matrix entries then VdistEq is defined to be the sum of log2 m - pi log2 pi / m where pi is the number of distance matrix entries equal to i.

❍ **vdw_area:** Area of van der Waals surface (A^2) calculated using a connection table approximation.

❍ **vdw_vol:** van der Waals volume (A^3) calculated using a connection table approximation.

❍ **vol:** van der Waals volume calculated using a grid approximation (spacing 0.75 A).

❍ **Vol_IMGT:** Volume value of amino acids, as reported by IMGT itself.

❍ **VSA:** van der Waals surface area. A polyhedral representation is used for each atom in calculating the surface area.

❍ **vsa_acc:** Approximation to the sum of VDW surface areas (A^2) of pure hydrogen bond acceptors (not counting acidic atoms and atoms that are both hydrogen bond donors and acceptors such as -OH).

❍ **vsa_acid:** Approximation to the sum of VDW surface areas of acidic atoms (A^2).

❍ **vsa_base:** Approximation to the sum of VDW surface areas of basic atoms (A^2).

❍ **vsa_don:** Approximation to the sum of VDW surface areas of pure hydrogen bond donors (not counting basic atoms and atoms that are both hydrogen bond donors and acceptors such as -OH) (A^2).

❍ **vsa_hyd:** Approximation to the sum of VDW surface areas of hydrophobic atoms (A^2).

❍ **vsa_other:** Approximation to the sum of VDW surface areas (A^2) of atoms typed as "other".

❍ **vsa_pol:** Approximation to the sum of VDW surface areas (A^2) of polar atoms (atoms that are both hydrogen bond donors and acceptors), such as -OH.

❍ **vsurf_A:** Amphiphilic moment.

❍ **vsurf_CP:** Critical packing parameter.

❍ **vsurf_CW1:** Capacity factor.

❍ **vsurf_CW2:** Capacity factor.

❍ **vsurf_CW3:** Capacity factor.

- ❍ **vsurf_CW4:** Capacity factor.

- ❍ **vsurf_CW5:** Capacity factor.

- ❍ **vsurf_CW6:** Capacity factor.

- ❍ **vsurf_CW7:** Capacity factor.

- ❍ **vsurf_CW8:** Capacity factor.

- ❍ **vsurf_D1:** Hydrophobic volume.

- ❍ **vsurf_D2:** Hydrophobic volume.

- ❍ **vsurf_D3:** Hydrophobic volume.

- ❍ **vsurf_D4:** Hydrophobic volume.

- ❍ **vsurf_D5:** Hydrophobic volume.

- ❍ **vsurf_D6:** Hydrophobic volume.

- ❍ **vsurf_D7:** Hydrophobic volume.

- ❍ **vsurf_D8:** Hydrophobic volume.

- ❍ **vsurf_DD12:** Contact distances of vsurf_DDmin.

- ❍ **vsurf_DD13:** Contact distances of vsurf_DDmin.

- ❍ **vsurf_DD23:** Contact distances of vsurf_DDmin.

- ❍ **vsurf_DW12:** Contact distances of vsurf_EWmin.

- **vsurf_DW13:** Contact distances of vsurf_EWmin.

- **vsurf_DW23:** Contact distances of vsurf_EWmin.

- **vsurf_EDmin1:** Lowest hydrophobic energy.

- **vsurf_EDmin2:** Lowest hydrophobic energy.

- **vsurf_EDmin3:** Lowest hydrophobic energy.

- **vsurf_EWmin1:** Lowest hydrophilic energy.

- **vsurf_EWmin2:** Lowest hydrophilic energy.

- **vsurf_EWmin3:** Lowest hydrophilic energy.

- **vsurf_G:** Surface globularity.

- **vsurf_HB1:** H-bond donor capacity.

- **vsurf_HB2:** H-bond donor capacity.

- **vsurf_HB3:** H-bond donor capacity.

- **vsurf_HB4:** H-bond donor capacity.

- **vsurf_HB5:** H-bond donor capacity.

- **vsurf_HB6:** H-bond donor capacity.

- **vsurf_HB7:** H-bond donor capacity.

- **vsurf_HB8:** H-bond donor capacity.

- **vsurf_HL1:** Hydrophilic-Lipophilic.

○ **vsurf_HL2:** Hydrophilic-Lipophilic.

○ **vsurf_ID1:** Hydrophobic integy moment.

○ **vsurf_ID2:** Hydrophobic integy moment.

○ **vsurf_ID3:** Hydrophobic integy moment.

○ **vsurf_ID4:** Hydrophobic integy moment.

○ **vsurf_ID5:** Hydrophobic integy moment.

○ **vsurf_ID6:** Hydrophobic integy moment.

○ **vsurf_ID7:** Hydrophobic integy moment.

○ **vsurf_ID8:** Hydrophobic integy moment.

○ **vsurf_IW1:** Hydrophilic integy moment.

○ **vsurf_IW2:** Hydrophilic integy moment.

○ **vsurf_IW3:** Hydrophilic integy moment.

○ **vsurf_IW4:** Hydrophilic integy moment.

○ **vsurf_IW5:** Hydrophilic integy moment.

○ **vsurf_IW6:** Hydrophilic integy moment.

○ **vsurf_IW7:** Hydrophilic integy moment.

○ **vsurf_IW8:** Hydrophilic integy moment.

❍ **vsurf_R:** Surface rugosity.

❍ **vsurf_S:** Interaction field surface area.

❍ **vsurf_V:** Interaction field volume.

❍ **vsurf_W1:** Hydrophilic volume.

❍ **vsurf_W2:** Hydrophilic volume.

❍ **vsurf_W3:** Hydrophilic volume.

❍ **vsurf_W4:** Hydrophilic volume.

❍ **vsurf_W5:** Hydrophilic volume.

❍ **vsurf_W6:** Hydrophilic volume.

❍ **vsurf_W7:** Hydrophilic volume.

❍ **vsurf_W8:** Hydrophilic volume.

❍ **vsurf_Wp1:** Polar volume.

❍ **vsurf_Wp2:** Polar volume.

❍ **vsurf_Wp3:** Polar volume.

❍ **vsurf_Wp4:** Polar volume.

❍ **vsurf_Wp5:** Polar volume.

❍ **vsurf_Wp6:** Polar volume.

❍ **vsurf_Wp7:** Polar volume.

❍ **vsurf_Wp8:** Polar volume.

❍ **Weight:** Molecular weight (including implicit hydrogens) in atomic mass units with atomic weights taken from CRC Handbook of Chemistry and Physics. CRC Press (1994).

❍ **wienerPath:** Wiener path number: half the sum of all the distance matrix entries as defined in [*Balaban, A.T.; Five New Topological Indices for the Branching of Tree-Like Graphs; Theoretica Chimica Acta 53 (1979) 355-375.*] and [*Wiener, H.; Structural Determination of Paraffin Boiling Points; Journal of the American Chemical Society 69 (1947) 17-20.*].

❍ **wienerPol:** Wiener polarity number: half the sum of all the distance matrix entries with a value of 3 as defined in Balaban, A.T.; Five New Topological Indices for the Branching of Tree-Like Graphs; Theoretica Chimica Acta 53 (1979) 355-375.

❍ **zagreb:** Zagreb index: the sum of d_i^2 over all heavy atoms i.

Reference:
Molecular Operating Environment (MOE), 2012.10;
Chemical Computing Group Inc.,
1010 Sherbooke St. West, Suite #910,
Montreal, QC, Canada, H3A 2R7, **2012**.