**Prediction of Mutagenicity of Chemicals from Their Calculated Molecular Descriptors: A Case Study with Structurally Homogeneous versus Diverse Datasets**

**Supplementary materials**

Table S1. Symbols, definitions and classification of topological indices

|  |  |  |  |
| --- | --- | --- | --- |
|  | Topostructural (TS) | In 508 compound data? | In 95  Amine data? |
| *IWD* | Information index for the magnitudes of distances between all possible pairs of vertices of a graph | **✓** | **✓** |
| *IWD* | Mean information index for the magnitude of distance | **✓** | **✓** |
| *W* | Wiener index = half-sum of the off-diagonal elements of the distance matrix of a graph | **✓** | **✓** |
| *ID* | Degree complexity | **✓** | **✓** |
| *HV* | Graph vertex complexity | **✓** | **✓** |
| *HD* | Graph distance complexity | **✓** | **✓** |
| *IC* | Information content of the distance matrix partitioned by frequency of occurrences of distance *h* | **✓** | **✓** |
| *M1* | A Zagreb group parameter = sum of square of degree over all vertices | **✓** | **✓** |
| *M2* | A Zagreb group parameter = sum of cross-product of degrees over all neighboring (connected) vertices | **✓** | **✓** |
| *hχ* | Path connectivity index of order *h* = 0-10 | **✓** | **✓** |
| *hχC* | Cluster connectivity index of order *h* = 3-5 | **✓** | **✓** |
| Cluster connectivity index of order *h* = *6* | **✓** | **✗** |
| *hχPC* | Path-cluster connectivity index of order *h* = 4-6 | **✓** | **✓** |
| *hχCh* | Chain connectivity index of order *h* = 3, 4, 7, 8 | **✓** | **✗** |
| Chain connectivity index of order *h* = 5, 6, 9, 10 | **✓** | **✓** |
| *Ph* | Number of paths of length *h* = 0-10 | **✓** | **✓** |
| *J* | Balaban’s *J* index based on topological distance | **✓** | **✓** |
| *nrings* | Number of rings in a graph | **✓** | **✓** |
| *ncirc* | Number of circuits in a graph | **✓** | **✓** |
| DN2S*y* | Triplet index from distance matrix, square of graph order, and distance sum; operation *y* = 1-4 | **✓** | **✓** |
| DN21*y* | Triplet index from distance matrix, square of graph order, and number 1; operation *y* = 1-5 | **✓** | **✓** |
| AS1*y* | Triplet index from adjacency matrix, distance sum, and number 1; operation *y* = 1-5 | **✓** | **✓** |
| DS1*y* | Triplet index from distance matrix, distance sum, and number 1; operation *y* = 1-2 | **✓** | **✓** |
| Triplet index from distance matrix, distance sum, and number 1; operation *y* = 3-5 | **✗** | **✓** |
| ASN*y* | Triplet index from adjacency matrix, distance sum, and graph order; operation *y* = 1-5 | **✓** | **✓** |
| DSN*y* | Triplet index from distance matrix, distance sum, and graph order; operation *y* = 1-5 | **✓** | **✓** |
| DN2N*y* | Triplet index from distance matrix, square of graph order, and graph order; operation *y* = 1-5 | **✓** | **✓** |
| ANS*y* | Triplet index from adjacency matrix, graph order, and distance sum; operation *y* = 1-2 | **✓** | **✓** |
| Triplet index from adjacency matrix, graph order, and distance sum; operation *y* = 3-5 | **✗** | **✓** |
| AN1*y* | Triplet index from adjacency matrix, graph order, and number 1; operation *y* = 1-5 | **✓** | **✓** |
| ANN*y* | Triplet index from adjacency matrix, graph order, and graph order again; operation *y* = 1-4 | **✓** | **✓** |
| Triplet index from adjacency matrix, graph order, and graph order again; operation *y* = 5 | **✓** | **✗** |
| ASV*y* | Triplet index from adjacency matrix, distance sum, and vertex degree; operation *y* = 1-2 | **✓** | **✓** |
| Triplet index from adjacency matrix, distance sum, and vertex degree; operation *y* = 3-5 | **✗** | **✓** |
| DSV*y* | Triplet index from distance matrix, distance sum, and vertex degree; operation *y* = 1-2 | **✓** | **✓** |
| ANV*y* | Triplet index from adjacency matrix, graph order, and vertex degree; operation *y* = 1-5 | **✓** | **✓** |
| *kp0* | Kappa zero | **✓** | **✓** |
| *kp1-kp3* | Kappa simple indices | **✓** | **✓** |
|  | **Topochemical (TC)** | **In 95 amine data?** | **In 508 compound data?** |
| O | Order of neighborhood when *ICr* reaches its maximum value for the hydrogen-filled graph | **✓** | **✓** |
| O*orb* | Order of neighborhood when *ICr* reaches its maximum value for the hydrogen-suppressed graph | **✓** | **✓** |
| I*ORB* | Information content or complexity of the hydrogen-suppressed graph at its maximum neighborhood of vertices | **✓** | **✓** |
| IC*r* | Mean information content or complexity of a graph based on the *r*th (*r* = 0-6) order neighborhood of vertices in a hydrogen-filled graph | **✓** | **✓** |
| SIC*r* | Structural information content for *r*th (*r* = 0-6) order neighborhood of vertices in a hydrogen-filled graph | **✓** | **✓** |
| CIC*r* | Complementary information content for *r*th (*r* = 0-6) order neighborhood of vertices in a hydrogen-filled graph | **✓** | **✓** |
| *hχb* | Bond path connectivity index of order *h* = 0-6 | **✓** | **✓** |
| *hχbC* | Bond cluster connectivity index of order *h* = 3, 5 | **✓** | **✓** |
| Bond cluster connectivity index of order *h* = 4, 6 | **✓** | **✗** |
| *hχbCh* | Bond chain connectivity index of order *h* = 4 | **✓** | **✗** |
| Bond chain connectivity index of order *h* = 5, 6 | **✓** | **✓** |
| *hχbPC* | Bond path-cluster connectivity index of order *h* = 4-6 | **✓** | **✓** |
| *hχv* | Valence path connectivity index of order *h* = 0-6 | **✓** | **✓** |
| *hχvC* | Valence cluster connectivity index of order *h* = 3, 5 | **✓** | **✓** |
| Valence cluster connectivity index of order *h* = 4, 6 | **✓** | **✗** |
| *hχvCh* | Valence chain connectivity index of order *h* = 3, 4, 7, 8 | **✓** | **✗** |
| Valence chain connectivity index of order *h* = 5, 6, 9, 10 | **✓** | **✓** |
| *hχvPC* | Valence path-cluster connectivity index of order *h* = 4-6 | **✓** | **✓** |
| *JB* | Balaban’s *J* index based on bond types | **✓** | **✓** |
| *JX* | Balaban’s *J* index based on relative electronegativities | **✓** | **✓** |
| *JY* | Balaban’s *J* index based on relative covalent radii | **✓** | **✓** |
| AZV*y* | Triplet index from adjacency matrix, atomic number, and vertex degree; operation *y* = 1-5 | **✓** | **✓** |
| AZS*y* | Triplet index from adjacency matrix, atomic number, and distance sum; operation *y* = 1-5 | **✓** | **✓** |
| ASZ*y* | Triplet index from adjacency matrix, distance sum, and atomic number; operation *y* = 1-2 | **✓** | **✓** |
| Triplet index from adjacency matrix, distance sum, and atomic number; operation *y* = 3-5 | **✗** | **✓** |
| AZN*y* | Triplet index from adjacency matrix, atomic number, and graph order; operation *y* = 1-5 | **✓** | **✓** |
| ANZ*y* | Triplet index from adjacency matrix, graph order, and atomic number; operation *y* = 1-2 | **✓** | **✓** |
| Triplet index from adjacency matrix, graph order, and atomic number; operation *y* = 3-5 | **✗** | **✓** |
| DSZ*y* | Triplet index from distance matrix, distance sum, and atomic number; operation *y* = 1,2 | **✓** | **✓** |
| DN2Z*y* | Triplet index from distance matrix, square of graph order, and atomic number; operation *y* = 1-2 | **✓** | **✓** |
| DN2Z*y* | Triplet index from distance matrix, square of graph order, and atomic number; operation 3-5 | **✗** | **✓** |
| *nvx* | Number of non-hydrogen atoms in a molecule | **✓** | **✓** |
| *nelem* | Number of elements in a molecule | **✓** | **✓** |
| *fw* | Molecular weight | **✓** | **✓** |
| *si* | Shannon information index | **✓** | **✓** |
| *totop* | Total Topological Index *t* | **✓** | **✓** |
| *sumI* | Sum of the intrinsic state values *I* | **✓** | **✓** |
| *sumdelI* | Sum of delta-*I* values | **✓** | **✓** |
| *tets2* | Total topological state index based on electrotopological state indices | **✓** | **✓** |
| *phia* | Flexibility index (*kp*1\* *kp*2/*nvx*) | **✓** | **✓** |
| *Idcbar* | Bonchev-Trinajstić information index | **✓** | **✓** |
| *IdC* | Bonchev-Trinajstić information index | **✓** | **✓** |
| *Wp* | Wiener *p* | **✓** | ✓ |
| *Pf* | Platt *f* | **✓** | ✓ |
| *Wt* | Total Wiener number | **✓** | ✓ |
| *knotp* | Difference of chi-cluster-3 and path/cluster-4 | **✓** | ✓ |
| *knotpv* | Valence difference of chi-cluster-3 and path/cluster-4 | **✓** | ✓ |
| *nclass* | Number of classes of topologically (symmetry) equivalent graph vertices | **✓** | ✓ |
| *NumHBd* | Number of hydrogen bond donors | **✓** | ✓ |
| *NumHBa* | Number of hydrogen bond acceptors | **✓** | ✓ |
| *SHCsats* | E-State of C *sp3* bonded to other saturated C atoms | **✓** | ✓ |
| *SHCsatu* | E-State of C *sp3* bonded to unsaturated C atoms | **✓** | ✓ |
| *SHvin* | E-State of C atoms in the vinyl group, *=CH-* | **✓** | ✗ |
| *SHtvin* | E-State of C atoms in the terminal vinyl group, *=CH2* | **✓** | ✗ |
| *SHarom* | E-State of C *sp2* which are part of an aromatic system | **✓** | ✓ |
| *SHHBd* | Hydrogen bond donor index, sum of Hydrogen E-State values for *–OH*, *=NH*, -*NH2*, *-NH-,-SH*, and *#CH* | **✓** | ✓ |
| *SHwHBd* | Weak hydrogen bond donor index, sum of *C-H* Hydrogen E-State values for hydrogen atoms on a C to which a F and/or Cl are also bonded | **✓** | ✗ |
| *SHHBa* | Hydrogen bond acceptor index, sum of the *E*-State values for *–OH*, *=NH*, *-NH2*, -*NH-*, *>N*, *-O-*, *-S-*, along with –F and –Cl | **✓** | ✓ |
| *Qv* | General Polarity descriptor | **✓** | ✓ |
| *NHBinty* | Count of potential internal hydrogen bonders (*y* = 2) | **✓** | ✗ |
| Count of potential internal hydrogen bonders (*y* = 3-10) | **✓** | ✓ |
| *SHBinty* | E-State descriptors of potential internal hydrogen bond strength (*y =*2, 5-9) | **✓** | ✗ |
| E-State descriptors of potential internal hydrogen bond strength (*y =*3, 4) | **✓** | ✓ |
| *ka1-ka3* | Kappa alpha indices | **✓** | ✓ |
|  | Electrotopological State index values for atom types:  *SHdNH, SHsSH, HssNH, SHtCH, SHCHnX, Hmaxpos, Hminneg, SsLi, SssBe, Sssss, Bem, SssBH ,SsssB, SssssBm, SdCH2, StCH, SddC, StsC, SdssC, SaasC, SsNH3p, SssNH2p, SdNH, StN, SsssNHp, SdsN, SsssN, SaasN, SssssNp, SaaO, SsSiH3, SssSiH2, SsssSiH, SssssSi, SsPH2, SssPH, SsssP, SdsssP, SsssssP, SsSH, SdS, SaaS, SdssS, SddssS, SssssssS, SsGeH3, SssGeH2, SsssGeH, SssssGe, SsAsH2, SssAsH, SsssAs, SdsssAs, SsssssAs, SsSeH, SdSe, SssSe, SaaSe, SdssSe, SddssSe, SsSnH3, SssSnH2, SsssSnH, SssssSn, SsI, SsPbH3, SssPbH2, SsssPbH, SssssPb* | **✓** | **✗** |
|  | Electrotopological State index values for atom types:  *SHsOH, SHsNH2, SHssNH, SHother, Hmax, Gmax, Hmin, Gmin, SsCH3, SssCH2, SdsCH, SaaCH, SsssCH, SaaaC, SssssC, SsNH2, SssNH, SaaNH, SaaN, SddsN, SsOH, SdO, SssO, SsF, SssS, SsCl, SsBr,* | **✓** | **✓** |
|  | **Geometrical (3-D)** | **In 95 amine data?** | **In 508 compound data?** |
| *3DW* | 3D Wiener number based on the hydrogen-suppressed geometric distance matrix | **✓** | **✓** |
| *3DW H* | 3D Wiener number based on the hydrogen-filled geometric distance matrix | **✓** | **✓** |
| *VW* | Van der Waal’s volume | **✓** | **✓** |
|  | Quantum Chemical (QC) | **In 95 amine data?** | **In 508 compound data?** |
| *EHOMO* | Energy of the highest occupied molecular orbital | **✓** | **✓** |
| *EHOMO-1* | Energy of the second highest occupied molecular | **✓** | **✓** |
| *ELUMO* | Energy of the lowest unoccupied molecular orbital | **✓** | **✓** |
| *ELUMO+1* | Energy of the second lowest unoccupied molecular orbital | **✓** | **✓** |
| *ΔHf* | Heat of formation | **✓** | **✓** |
| *μ* | Dipole moment | **✓** | **✓** |