1. Constitutional indices (47)

|  |  |  |
| --- | --- | --- |
| Name | Description | 备注 |
| MW | molecular weight | 分子质量 |
| AMW | average molecular weight | 平均分子质量 |
| Sv | sum of atomic van der Waals volumes | 原子范德华体积之和 |
| Se | sum of atomic Sanderson electronegativities | 原子电负性之和 |
| Sp | sum of atomic polarizabilities | 原子极性之和 |
| Si | sum of first ionization potentials | 第一电离能之和 |
| Mv | mean atomic van der Waals volume | 平均原子范德华体积 |
| Me | mean atomic Sanderson electronegativity | 平均原子电负性 |
| Mp | mean atomic polarizability | 平均原子极性 |
| Mi | mean first ionization potential | 平均电离能 |
| GD | graph density | 图像密度 |
| nAT | number of atoms | 原子个数 |
| nSK | number of non-H atoms | 非H原子个数（顶点个数） |
| nTA | number of terminal atoms | 只有一根非H键的原子数 |
| nBT | number of bonds | 键数 |
| nBO | number of non-H bonds | 不含H键数 |
| nBM | number of multiple bonds | 多重键数 |
| SCBO | sum of conventional bond orders | 传统键级之和，单1双2三3芳香1.5去H |
| RBN | number of rotatable bonds | 可旋转键数（不包括C-N） |
| RBF | rotatable bond fraction | 可旋转键占比 |
| nDB | number of double bonds | 双键 |
| nTB | number of triple bonds | 三键 |
| nAB | number of aromatic bonds | 芳香键 |
| n ‘i’ | number of ‘I’ atoms | i为 H,C,N,O,P,S,F,Cl,Br,I,B |
| nHM | number of heavy atoms | 重原子数 |
| nHet | number of heteroatoms | 杂原子数（非H原子数） |
| nX | number of halogen atoms | 卤素原子数 |
| ‘i’% | percentage of ‘I’ atoms | i为HCNO,以及halogen |
| nCsp3 | number of sp3 hybridized Carbon atoms | SP3 |
| nCsp2 | number of sp2 hybridized Carbon atoms | SP2 |
| nCsp | number of sp hybridized Carbon atoms | SP |
| nStructure | number of disconnected structures | 没有成键的结构 |
| totalcharge | total charge | 总电荷 |

1. Ring descriptors (32)

|  |  |  |
| --- | --- | --- |
| Name | Description | 备注 |
| nCIC | number of rings | rings的单元可以重复，circuit不可以重复 |
| nCIR | number of circuits |
| TRS | total ring size | 独立环求和（即删除了一些联环） |
| Reprim | ring perimeter | 单环数\*2 – TRS |
| Rbird | ring bridge count | 桥环数量 |
| MCD | molecular cyclized degree | 分子环化程度 |
| RFD | ring fusion density | 环融合密度 |
| RCI | ring complexity index | 环复杂程度 |
| NRS | number of rings systems | ring系统个数 |
| NNRS | normalized number of ring systems | 归一化的ring系统个数 |
| nRi, i=range(3,12) | number of 3~12-membered rings | 3~12原子的ring个数 |
| nBnz | number of benzene-like rings | 苯环数 |
| ARR | aromatic ratio | 芳香比 |
| D/Dtri, i=range(3,12) | distance/detour ring index of order 3~12 |  |

1. Topological indices (75)

分为6类，Vertex degree-based indices（顶点指数），Distance-based indices（距离指数），MTI（molecular topological index）（分子拓扑指数），Path/walk indices（路径指数），E-state indices（电子拓扑态指数），Centric indices（中心指数）。

**Vertex degree-based indices**

Vertex degree衍生自molecular graph local quantity。每个原子的vertex degree就是对应邻接矩阵的列和。

|  |  |  |
| --- | --- | --- |
| Name | Description | 备注 |
| ZM1 | first Zagreb index | vertex degree的平方和（不算H） |
| ZM2 | second Zagreb index | vertex degree的平方和（算H） |
| ZM1V | first Zagreb index by valence vertex degrees | 价电子邻接矩阵生成的vertex degree的平方和（不算H） |
| ZM2V | second Zagreb index by valence vertex degrees | 价电子邻接矩阵生成的vertex degree的平方和（算H） |
| ZMkV | kth Zagreb index by Madan vertex degrees | 加上Madan化学度权重的邻接矩阵生成的vertex degree的平方和 |
| ZMkPer | kth Zagreb index by perturbation vertex degrees | perturbation 权重 |

**Distance-based indices**

基于拓扑距离的描述符。

拓扑距离：两个原子间最短的path

|  |  |  |
| --- | --- | --- |
| Name | Description | 备注 |
| PoI | polarity number | 拓扑距离3为半径范围内的极性顶点对的个数 |
| MSD | mean square distance index | 平均均方距离 |
| SPI | superpendentic index | 直译为‘超依赖指数’，reduced distance matrix中每列数字和的均方根 |
| PJI2 | 2D Petitjean shape index | 拓扑直径-半径与半径比。拓扑直径定义为原子最大偏心距，半径定位最小偏心距。 |
| ECC | eccentricity | 偏心距和 |
| AECC | average eccentricity | 平均偏心距 |
| MDDD | mean distance degree deviation | 平均距离度偏差 |
| UNIP | unipolarity | 最小顶点距离度 |
| CENT | centralization | 2Wi\_D – nSK UNIP，Wi\_D是Wiener index |
| VAR | variation |  |
| ICR | radial centric information index | 径向中心信息指数    nk是相同原子偏心距的顶点的个数 |

**molecular topological index**

计算自邻接矩阵A与距离矩阵D。

|  |  |  |
| --- | --- | --- |
| Name | Description | 备注 |
| SMTI | Schultz index | v是列向量 |
| SMTIV |  | 与SMTI一样，但是将顶点度换成了价层顶点度 |
| GMTI | GMTI index | 是顶点度，dij是拓扑距离 |
| GMTIV |  | 顶点度换成了价层顶点度 |
| Xu | Xu index | 是顶点度，是顶点距离度 |
| CSI | eccentric connectivity index | 非H原子偏心距、顶点度之和 |

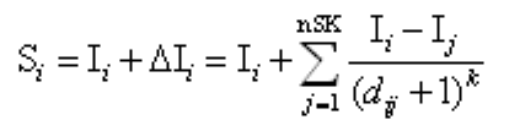
**Path/walk indices**

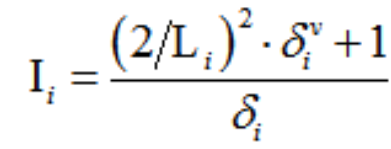
分子中一个原子到相邻成键原子的“步行”为一个walk。在对分子计数walk时，任何顶点可以被任意次到达。与之相对应的是path，即在对分子计数path时，每个顶点只能被到达一次。walk length 和path length就是对应的walk与path长度。walk degree 和path degree是对应所有顶点出发计数walk或path的总和。

|  |  |  |
| --- | --- | --- |
| Name | Description | 备注 |
| Wap | all-path Wiener index | 分子非H部分的path degree的一半 |
| S1K | 1 Kier alpha-modified shape index | 与分子复杂度有关 |
| S2K | 2 Kier alpha-modified shape index | 分子中原子空间密度的信息 |
| S3K | 3 Kier alpha-modified shape index | 与支链的中心性有关 |
| PHI | kier flexibility index |  |
| PWk | path/walk k – Randic shape index | k级path个数\*nSK/k级walk个数。  k级path个数指的是，path为1的个数。 |

**E-state indices**

首先定义电子拓扑态指数 electrotopological state indices:



I是intrinsic state，；I是其他原子引起的场效应。

|  |  |  |
| --- | --- | --- |
| Name | Description | 备注 |
| MAXDN, MAXDP | maximal electrotopological negative/positive variation | 最大电拓扑负/正偏差  MAXDN是分子里最负的  MAXDP是分子里最正的 |
| DELS | molecular electrotopological variation | 分子电拓扑偏差，MAXDN-MAXDP |
| TIE | E-state topological parameter | 电子态拓扑指数    nBO是非H键长数，nCIC是独立环数，S是电子拓扑态指数。 |
| Psi\_\*\_\* | Psi indexes | Psi类指数 |

**Centric indices**

|  |  |  |
| --- | --- | --- |
| Name | Description | 备注 |
| BAC | Balaban centric index |  |
| LOC | lopping centric index |  |

1. Walk and path counts (46)

|  |  |  |
| --- | --- | --- |
| Name | Description | 备注 |
| MWCi, i in range(1,10) | molecular walk count of order 1~10 |  |
| SRWi, i in range(1,10) | self-returning walk count of order 1~10 |  |
| MPCi, i in range(1,10) | molecular path count of order 1~10 |  |
| piPCi, i in range(1,10) | molecular multiple path count of order 1~10 |  |
| TWC | total walk count |  |
| TPC | total path count |  |
| pilD | conventional bond order ID number |  |
| PCR | ratio of multiple path count over path count |  |
| CID | Randic ID number |  |
| BID | Balaban ID number |  |

1. Connectivity indices (37)

用于描述分子中（不考虑H）被连接到的原子的个数。

|  |  |  |
| --- | --- | --- |
| Name | Description | 备注 |
| Xi, i in range(0,5) | connectivity index of order 0~5 | 连接指数 |
| XiA, i in range(0,5) | average connectivity index of order 0~5 | 平均连接指数 |
| Xiv, i in range(0,5) | valence connectivity index of order 0~5 |  |
| XiAv, i in range(0,5) | average valence connectivity index of order 0~5 |  |
| Xisol, i in range(0,5) | solvation connectivity index of order 0~5 |  |
| XMOD | modified Randic index |  |
| RDCHI | reciprocal distance sum Randic-like index |  |
| RDSQ | reciprocal distance sum inverse Randic-like index |  |
| X1Kup | Kupchik connectivity index |  |
| X1Mad | connectivity topochemical index |  |
| X1Per | perturbation connectivity index |  |
| X1MulPer | multiplicative perturbation connectivity index |  |

1. Information indices (50)

通过对元素成分进行抽象化（抽象成数字），来描述化学组成、成键方式、分子拓扑、对称性。

|  |  |  |
| --- | --- | --- |
| Name | Description | 备注 |
| ISIZ | information index on molecular size | 分子信息指数    将分子的原子分成几类等价原子，G为类的个数，ng是类里原子的个数 |
| IAC | total information index on atomic composition | 元素组成的总信息指数 |
| AAC | mean information index on atomic composition | 元素组成的平均信息指数 |
| IDET | total information content on the distance equality |  |
| IDE | mean information content on the distance equality |  |
| IDM | mean information content on the distance magnitude |  |
| IDDE | mean information content on the distance degree magnitude |  |
| IDET | total information content on the distance equality |  |
| IDMT | total information content on the distance magnitude |  |
| IVDE | mean information content on the vertex degree equality |  |
| IVDM | mean information content on the vertex degree magnitude |  |
| Ges | Number of symmetry classes (based on electrotopological state) |  |
| rGes | Relatibe number of symmetry classes (based on electrotopological state) |  |
| S0K | Kier symmetry index |  |
| HVcpx | graph vertex complexity index |  |
| HDcpx | graph distance complexity index |  |
| Uindex | Balaban U index |  |
| Vindex | Balaban V index |  |
| Xindex | Balaban X index |  |
| Yindex | Balaban Y index |  |
| pICi, p in list(‘ ’,T,S,C,B), i in range(0,5) | information content index | 20个 |

1. 2D matrix-based descriptors (607)

对9类graph-theoretical矩阵用代数算子计算。9类graph-theoretical为邻接矩阵，拓扑距离矩阵，拉普拉斯矩阵，Chi矩阵，倒数平方距离矩阵，迂回矩阵，距离/迂回矩阵，Barysz矩阵，Burden矩阵。

部分举例：

|  |  |  |
| --- | --- | --- |
| Name | Description | 备注 |
| J\_A | Balaban-like index from adjacency matrix | 基于邻接矩阵计算的类Balaban指数 |
| SpPos\_A | spectral positive sum from adjacency matrix | 邻接矩阵的光谱正向和 |
| SpPosA\_A | normalized spectral positive sum from adjacency matrix | 归一化的邻接矩阵的光谱正向和 |
| SpPosLog\_A | logarithmic spectral positive sum from adjacency matrix | log化的邻接矩阵的光谱正向和 |
| SpMax\_A | leading eigenvalue from adjacency matrix (Lovasz-Pelikan index) | 邻接矩阵的主要特征值 |

1. 2D autocorrelations (213)

部分举例：

|  |  |  |
| --- | --- | --- |
| Name | Description | 备注 |
| ATSim, i in range(1,8) | Broto-Moreau autocorrelation of lag 1~8 (log function) weighted by mass |  |
| ATSiv, i in range(1,8) | Broto-Moreau autocorrelation of lag 1~8 (log function) weighted by van der Waals volumes |  |
| ATSie, i in range(1,8) | Broto-Moreau autocorrelation of lag 1~8 (log function) weighted by Sanderson electronegativity |  |

1. Burden eigenvalues (96)

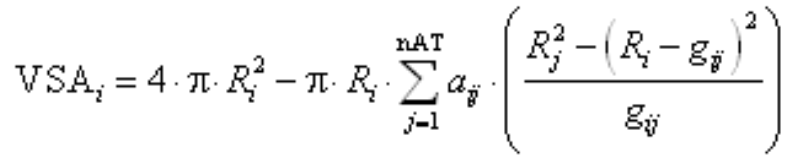
计算Burden矩阵的特征后的描述符

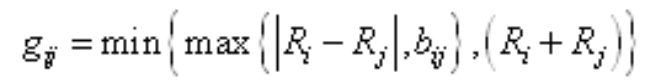
举例：

|  |  |  |
| --- | --- | --- |
| Name | Description | 备注 |
| SpMax1\_Bh(m) | largest eigenvalue n. 1 of Burden matrix weighted by mass | mass加权后，最大的特征值 |
| SpMax2\_Bh(m) | largest eigenvalue n. 2 of Burden matrix weighted by mass | 次大的特征值 |
| SpMax2\_Bh(v) | largest eigenvalue n. 2 of Burden matrix weighted by van der Waals volume | 范德华体积加权后的最大特征值 |

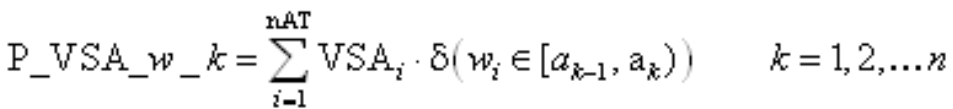
1. P\_VSA-like descriptor (55)

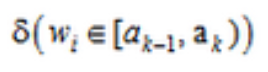
P. Labute 发明的van der Waals surface area (VSA) 类描述符。 i原子的VSA描述符可以被定义为：

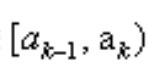
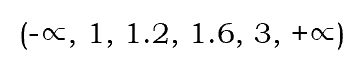


R定义为原子的范德华半径，是i与j原子的邻接矩阵。g为，R是ij原子的理想键长，r为相对键长，c是校正项，单键为0，芳香键为0.1，双键为0.2，三键为0.3。每个原子对的相对键长内容可以从help手册中查到。

如果我们考虑原子的某类性质x（质量、体积），那么关于性质x的P\_VSA类描述符可以被定义成：



是Dirac delta函数。

根据计算的性质不同，的范围也不同。如logP，那么则为中的一个，mass则为中的一个。

那么根据计算性质以及的不同，可以分为不同的描述符。

举例：

|  |  |  |
| --- | --- | --- |
| Name | Description | 备注 |
| P\_VSA\_LogP\_1 | P\_VSA-like on LogP, bin1 | 关于LogP的P\_VSA-like描述符，区间为,为第一个间隔，所以为bin1。 |
| P\_VSA\_m\_3 | P\_VSA-like on mass, bin 3 | 关于mass的P\_VSA-like描述符，区间为第三个的 |

此外P\_VSA-like描述符还基于Potential Pharmacophore Points (PPP)开发了药物方面的P\_VSA\_ppp\_x描述符。

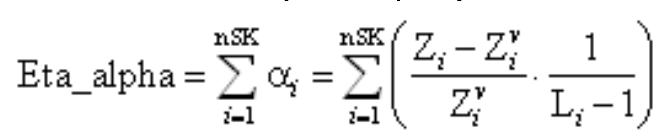
举例：

|  |  |  |
| --- | --- | --- |
| Name | Description | 备注 |
| P\_VSA\_ppp\_P | P\_VSA-like on potential pharmacophore points, P - positive |  |

1. Extended Topochemical Atom indices (ETA indices) (23)

主要是关于分子中非H原子部分的描述，分为价电子部分描述和原子核描述部分。

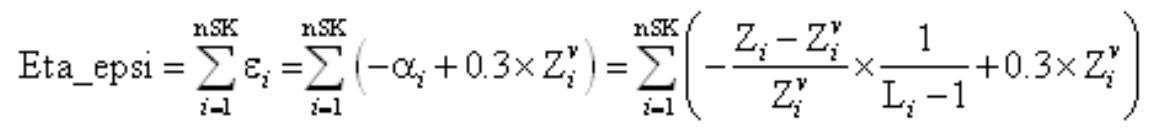
Eta\_alpha(core count)用来描述非价电子部分：



Z、Zv、L分别是原子数、价电子数、主量子数。

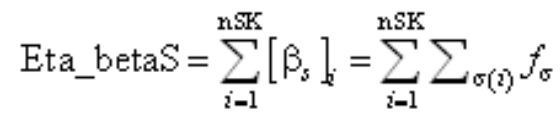
Eta\_alpha\_A是Eta\_alpha的平均（除以nSK）

Eta\_epsi用来描述电负性：

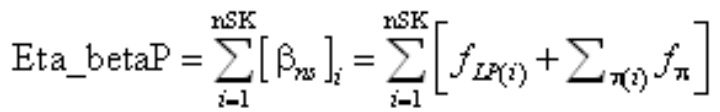


Eta\_epsi\_A是Eta\_epsi的平均（除以nSK）

Eta\_betaS是sigma VEM(valence electron mobile) count，应该计数sigma价电子数目：



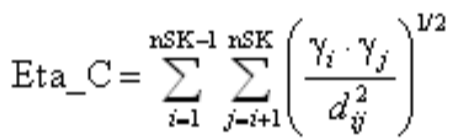
Eta\_betaP对应计算Pi键和大Pi键电子数：



Eta\_beta可以简单理解为上述两者之和。

Eta\_beta\_A为平均Eta\_beta

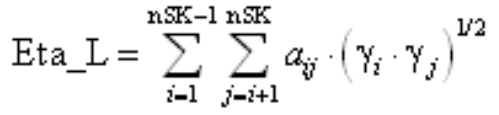
Eta\_C为composite ETA index，混合ETA指数：



其中，拓扑距离。

Eta\_C\_A为平均

Eta\_L为local composite index，局域混合ETA指数：



Eta\_L\_A为平均

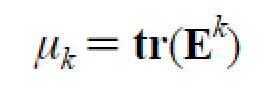
Eta\_F（functionality ETA index）、Eta\_F\_A、Eta\_FL（local functionality ETA index）是为了评估分子中功能性原子（杂原子、双三键）。

此外还有Eta\_B用来评估支链、Eta\_sh\_x用来评估x型（四键）基团、Eta\_sh\_y用来评估y型（三键）基团、Eta\_sh\_p用来评估pi基团。

1. Edge adjacency indices (324)

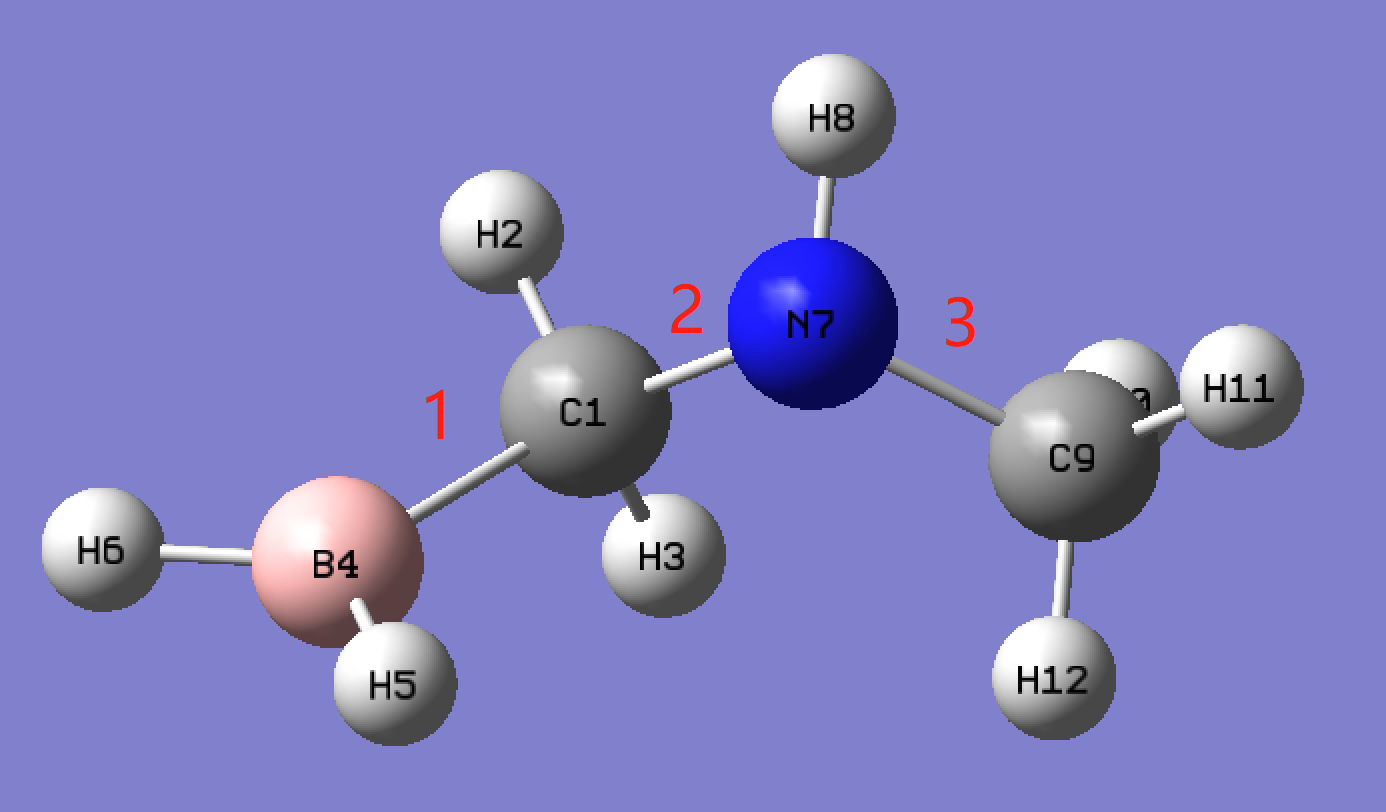
利用边缘邻接矩阵对spectral、connectivity-like、spectral moments、eigenvalues等拓扑指数或性质计算的描述符。

**关于spectral moments：**



E即边缘邻接矩阵。SK*k\_*M是每一级边缘邻接矩阵的迹。k级定义为k个path。

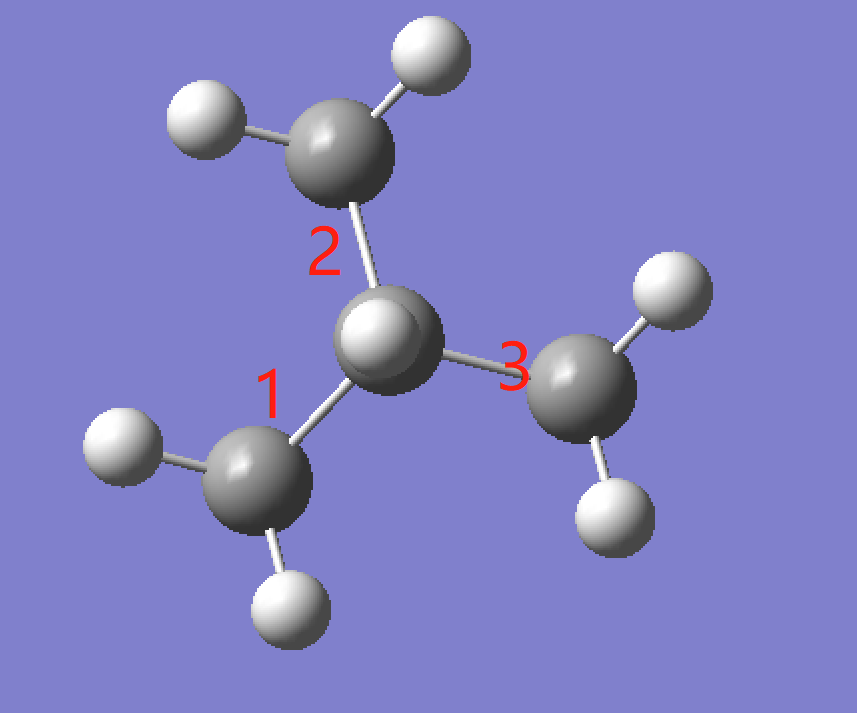
比如，B-C-N-C，定义三根键分别为123。



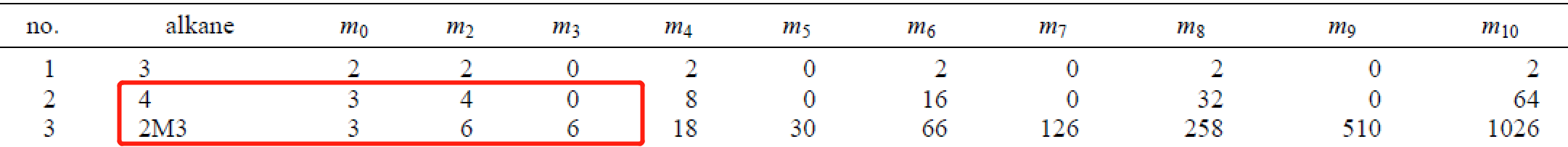
那么k=0,1,2,3,4（path=0,1,2,3,4）的边缘邻接矩阵为：

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| k=0 | 1 | 2 | 3 | k=1 | 1 | 2 | 3 | k=2 | 1 | 2 | 3 | k=3 | 1 | 2 | 3 | k=4 | 1 | 2 | 3 |
| 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 2 | 0 | 1 | 2 | 1 | 2 |
| 2 | 0 | 1 | 0 | 2 | 1 | 0 | 1 | 2 | 0 | 2 | 0 | 2 | 2 | 0 | 2 | 2 | 1 | 4 | 1 |
| 3 | 0 | 0 | 1 | 3 | 0 | 1 | 0 | 3 | 1 | 0 | 1 | 3 | 0 | 2 | 0 | 3 | 2 | 1 | 2 |
| trace | 3 |  |  | trace | 0 |  |  | trace | 4 |  |  | trace | 0 |  |  | trace | 8 |  |  |

对于异丁烷：



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| k=0 | 1 | 2 | 3 | k=1 | 1 | 2 | 3 | k=2 | 1 | 2 | 3 | k=3 | 1 | 2 | 3 |
| 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 2 | 2 | 2 |
| 2 | 0 | 1 | 0 | 2 | 1 | 0 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 2 |
| 3 | 0 | 0 | 1 | 3 | 1 | 1 | 0 | 3 | 1 | 1 | 2 | 3 | 2 | 2 | 2 |
| trace | 3 |  |  | trace | 0 |  |  | trace | 6 |  |  | trace | 6 |  |  |



**关于eigenvalues：**

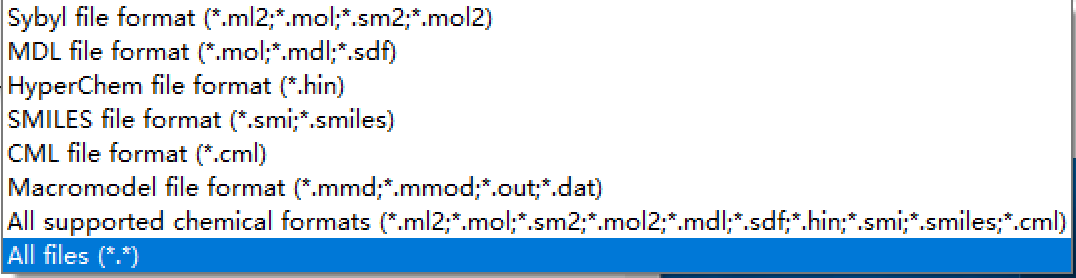
k=1时边缘邻接矩阵的特征值。如上述第一个分子的为1.4，0，-1.4；异丁烷的为2，-1，-1。

举例：

|  |  |  |
| --- | --- | --- |
| Name | Description | 备注 |
| SpMax\_EA | leading eigenvalue from edge adjacency mat | 边缘邻接矩阵的主要特征值 |
| SpDiam\_EA | spectral diameter from edge adjacency mat | 边缘邻接矩阵的光谱直径 |
| SpMax\_EA(dm) | leading eigenvalue from edge adjacency mat. weighted by dipole moment | 经由偶极矩加权后边缘邻接矩阵的主要特征值 |

1. Geometrical descriptors (38)

根据输入的原子坐标信息来得到关于分子3D结构描述符。值得一提的是，help手册中特别提到了也可以导入晶体坐标。关于导入的结构，可以检查一下下面的截图里是否有兼容的晶体文件。若没有的话，那还是只能导入去周期性的晶体文件。



“Geometrical descriptors are defined in several different ways but always derived from the three-dimensional structure of the molecule. Generally, geometrical descriptors are calculated either from some optimised molecular geometry obtained by the methods of the computational chemistry or **from crystallographic coordinates**.”

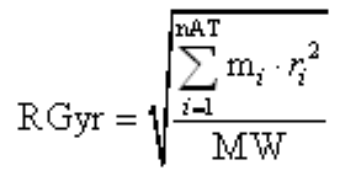
几何描述符可以分为四大类：**Size indices, Shape indices, Delocalization-degree indices, COMMA descriptors.**

**Size indices:**

G1 和 G2 都是 Gravitational indices，反应了分子的质量贡献，分别可以定义为：



RGyr（radius of gyration）是回旋半径，用来描述分子中每个原子的质量贡献：

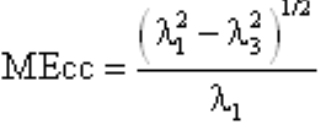


SPAN（span R）用来定义最小原子核的半径。

**Shape indices:**

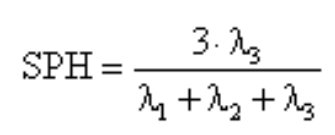
SPAM是SPAN的均方根。

MEcc（molecular eccentricity）描述的是分子离心率，定义为：

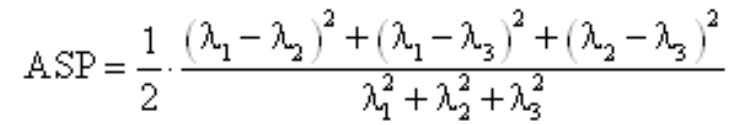


lambda为分子转动惯量矩阵的特征值。

SPH（spherosity）是一个椭圆描述符，计算子分子协方差矩阵的特征值：



ASP（asphericity）描述的是椭圆球形状的偏差：



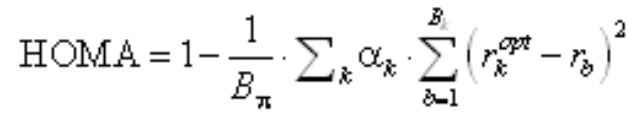
PJI3（3D Petitjean shape index）定义为最大几何离心度与最小几何离心度的差值比上最小几何离心度的商。

L/Bw定义为长宽比。

**Delocalization-degree indices**

广域尺度的描述符主要描述的是Pi电子有关的内容。

HOMA（Harmonic Oscillator Model of Aromaticity index）描述的是单双键相对于芳香键的偏差度：

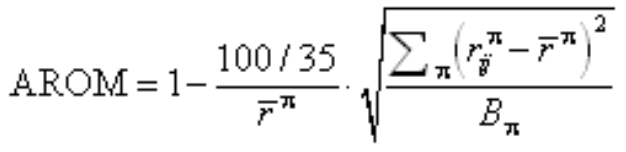


Bk是第k个键的Pi键个数，rb是真实键长，和是数值常数与常见的成键长度（help里有表）。

HOMT是HOMA×。

CMBL（conjugated maximum bond length）是成键系统中最弱键的键长。

AROM（Aromaticity）衍生于前述的芳香描述符：



**COMMA descriptors**

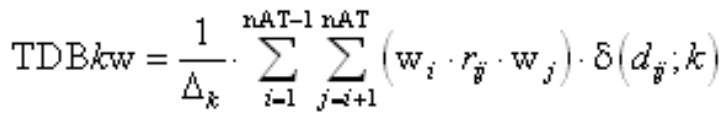
主要是几何结构（三维以及三个坐标轴方向）中心与质量中心、体积中心、极性中心等的偏差。

1. 3D matrix-based descriptors (99)

与2D matrix-based descriptors一样，只是维度不同。

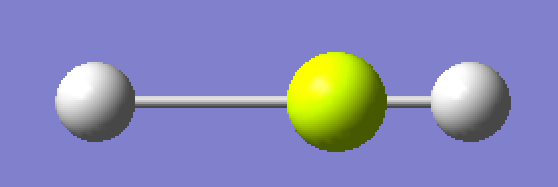
1. 3D Autocorrelations (80)

3D Autocorrelations主要描述的是原子间原子性质的相关性。通式为：



为拓扑距离在时的原子对数，w为原子性质，rij为原子对之间的几何距离，dij为拓扑距离，为binary值，当dij等于k时为1，否则为0。

以BH2为例子，设定一根键为1.8，一根为1.0。左H为H1，右H为H2。



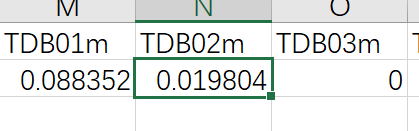
根据公式计算原子性质为mass且k=1时的TDB，此时TDB记为TDB1m，BH2的计算公式为：

上式中，拓扑距离为1的原子对有2个，因而为2，质量用实际质量带入。

同理有k=2时的TDB2m为：

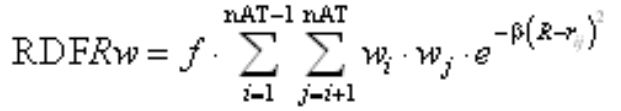
同理k=3时的TDB3m由于不存在拓扑距离为3的原子对，因而TDB3m为0。

与Dragon的结果一致：



1. RDF descriptors (210)

RDF（Radial Distribution Function）径向分布函数描述符基于径向分布函数计算得到，这里可以理解为在半径R体积圆上找到一个原子的概率。



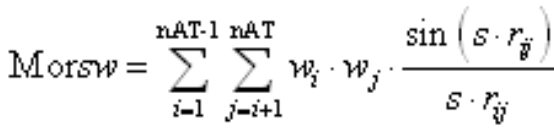
f是比例因子，是原子ij的性质（质量、体积等），rij是原子间距，是平滑参数（可以理解为温度因子），1个被定义为100A-2，1个R定义为0.5A。

举例：

|  |  |  |
| --- | --- | --- |
| Name | Description | 备注 |
| RDF010u | Radial Distribution Function – 010 / unweighted | 无加权的010单位R的RDF描述符 |
| RDF065m | Radial Distribution Function – 065 / weighted by mass | 质量加权的065单位R的RDF描述符 |

1. 3D-MoRSE descriptors (224)

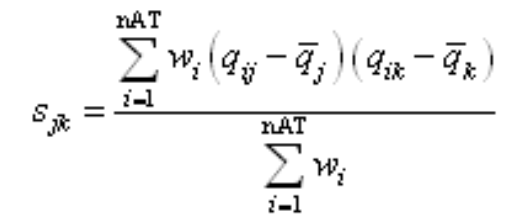
3D-Molecule Representation of Structures based on Electron diffraction，与RDF类似，3D-MoRSE descriptors是基于电子衍射学产生的描述符：



代表所有原子在不同方向上的衍射比，在dragon里被划分0~31个等级，0表示没有衍射。

1. WHIM descriptors (114)

Weighted Holistic Invariant Molecular descriptors，加权整体不变分子描述符，计算自原子在坐标轴上的投影。具体的算法是计算加权协方差矩阵的特征值与特征向量：

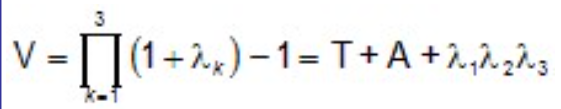


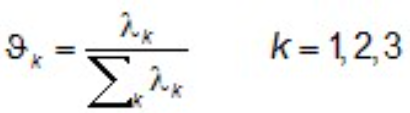
是ij之间的加权协方数，和代表原子坐标以及其平均坐标。计算所有的后得到协方差矩阵。

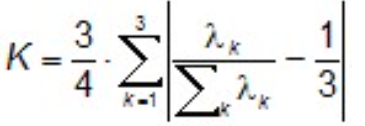
Lkw，k=1，2，3。指的是三个特征值。

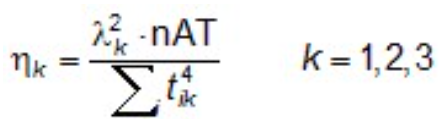
Tw，指的是三个特征值和。

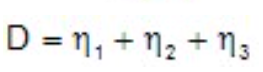
Aw，指的是

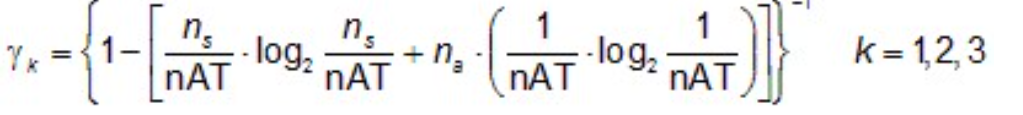
Vw，指的是

Pkw，描述轴上的形状：

Kw，描述整体形状：

EKw，描述轴上的密度：

Dw，整体密度：

Gkw，轴对称性：

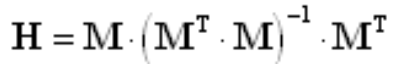
Gw，整体对称性

以上，特征值，t为原子每个坐标轴坐标，是一个每个坐标轴上对称原子的个数，是整体对称原子的个数。

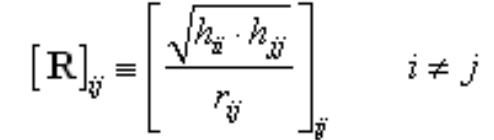
1. GATEWAY descriptor (273)

GEometry, Topology, and Atom-Weights AssemblY，是一类化学结构信息描述符。

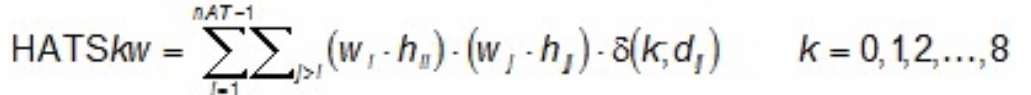
首先定义一个分子影响矩阵H（Molecular Influence Matrix）:

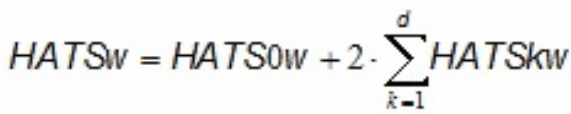


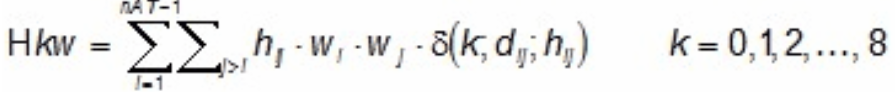
矩阵M是分子坐标填充的矩阵。定义影响/距离矩阵R(influence/distance matrix)为：

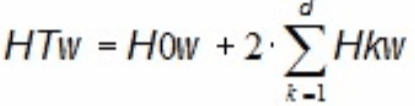


**H indices**

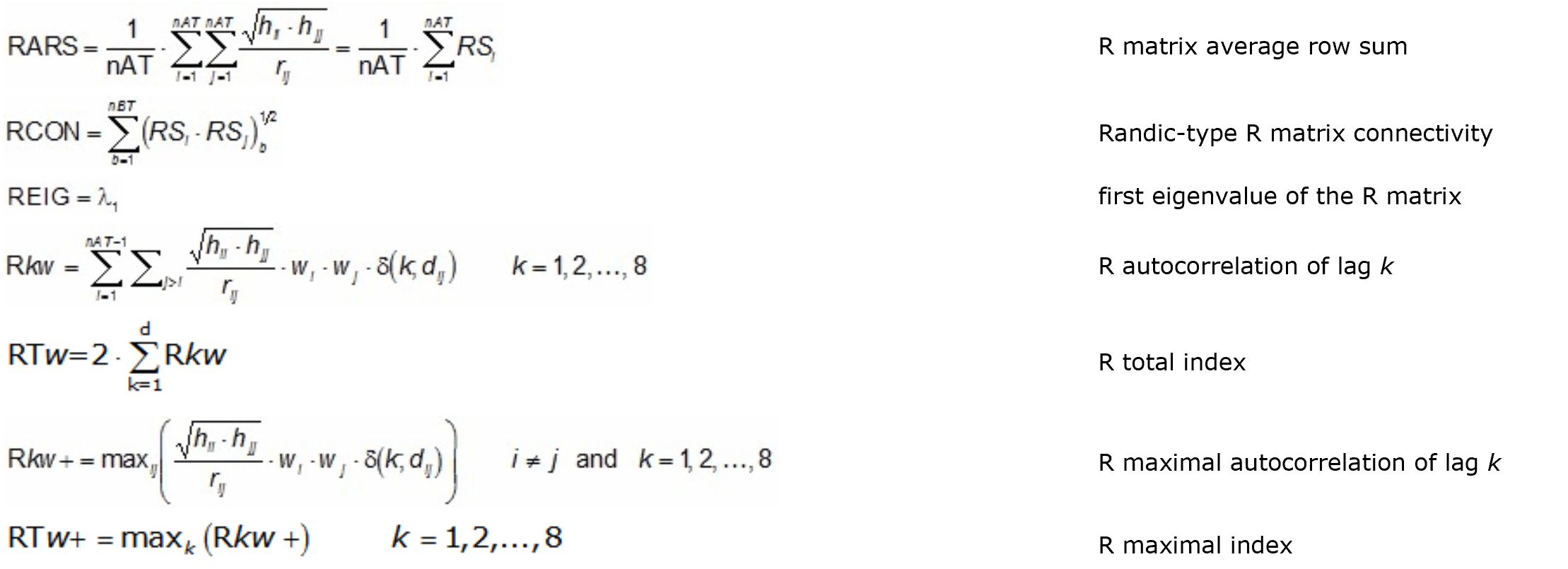
加权杠杆自相关描述符leverage-weighted autocorrelation：

加权杠杆总自相关描述符：

H矩阵自相关：

H矩阵总自相关：

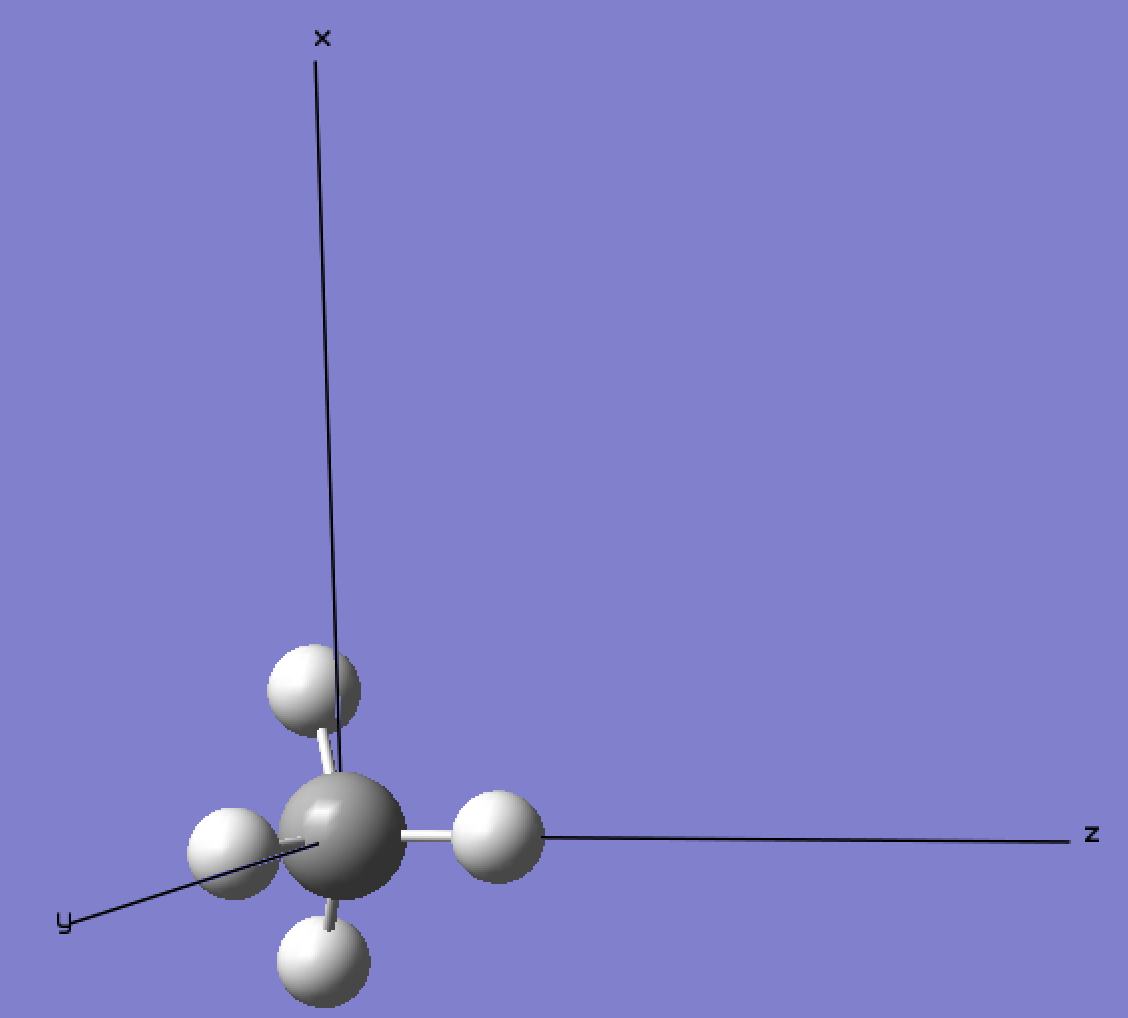
R indices与H indices类似，不累述。



举例：

关于Rkw+，例子用CH4分子，计算R1m+与R2m+。

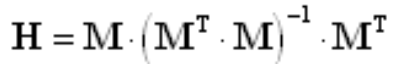
首先利用GV画出CH4分子，并将其处于坐标中心：



列出分子坐标矩阵M：

|  |  |  |  |
| --- | --- | --- | --- |
| 矩阵M | x | y | z |
| C | 0 | 0 | 0 |
| H | 0 | 0 | 1.07 |
| H | 0 | 1.0088 | -0.3567 |
| H | 0.8737 | -0.5044 | -0.3567 |
| H | -0.8737 | -0.5044 | -0.3567 |

根据H矩阵定义计算H矩阵：



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 矩阵H | C | H | H | H | H |
| C | 0 | 0 | 0 | 0 | 0 |
| H | 0 | 0.749965 | -0.250012 | -0.250012 | -0.250012 |
| H | 0 | -0.250012 | 0.750012 | -0.249988 | -0.249988 |
| H | 0 | -0.250012 | -0.249988 | 0.750012 | -0.249988 |
| H | 0 | -0.250012 | -0.249988 | -0.249988 | 0.750012 |

根据Rkw+公式计算R1m+与R2m+。k=1代表一个拓扑距离，在这里代表C-H原子对。四个C-H都是一样的，因而R1m中最大的R1m+就是4个中的任意一个，即：

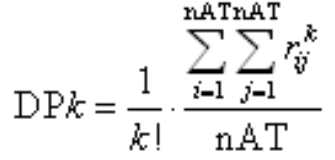
k=2代表两个拓扑距离，在这里代表H与H。四个H-H原子对都是一样的，因而同理：

对比dragon结果：



1. Randic molecular profiles (41)

Randic提出的一系列描述符，用来描述分子内原子间的几何距离：



特别适合用于分析分子的相似性与差异性。

1. Functional group counts (154)

官能团计数，即简单的对某个官能团计数。

1. Atom-centred fragments (115)

成键中心原子的计数，即人为地将某个成键碳定义成一类原子碳，比如CH3R定义为C-001，有多少个CH3R就计多少数。

1. Atom-type E-state indices (172)

原子型电子态描述符，混合了结构信息与电子信息。其实与21类似，只是这里多考虑了电荷信息。比如21中考虑-NH3基团，而这里会考虑-NH3[+1]。

1. CATS 2D (150)

Chemically Advanced Template Search，药物位点类的描述符。

1. 2D Atom Pairs (1596)

描述的分子中两个原子之间的信息。原子选取范围为是CNOSPFClBrIBSi以及X（X代表除了前面以外的任意原子）。信息选择有三个，拓扑距离、一定距离范围内是否存在、一定距离范围内出现次数。距离划分10个等级，1埃~10埃。

例子：

1. T(N…O) N原子与O原子的拓扑距离
2. B01[C-C] C原子周围是否存在C原子，周围拓扑半径为1埃，存在为1，不存在为0。B意思是binary，即二分问题，存在于不存在
3. B01[C-X] C原子周围是否存在其他任意原子，周围拓扑半径为1埃，存在为1，不存在为0。
4. F01[Si-C] Si原子周围出现C原子的次数，周围拓扑半径为1埃。F代表Frequency。
5. 3D Atom Pairs (36)

与25一样，只是判断距离时将拓扑半径换成了几何半径。

1. Charge descriptors (15)

电荷描述符

|  |  |  |
| --- | --- | --- |
| Name | Description | 备注 |
| qpmax | maximum positive charge | 最大正电荷 |
| qnmax | maximum negative charge | 最大负电荷 |
| Qpos | total positive charge | 总正电荷 |
| Qneg | total negative charge | 总负电荷 |
| Qtot | total absolute charge (electronic charge index - ECI) | 总绝对电荷（正电荷+负电荷的绝对值） |
| Qmean | mean absolute charge (charge polarization) | 总绝对电荷除以原子数 |
| Q2 | total squared charge | 原子电荷平方和 |
| RPCG | relative positive charge | 相对正电荷=最正原子电荷/总正电荷 |
| RNCG | relative negative charge | 相对负电荷 |
| SPP | submolecular polarity parameter | 同类元素原子最大差异电荷 |
| TE1 | topographic electronic descriptor | 局域电荷（成键原子与非成键原子都count） |
| TE2 | topographic electronic descriptor (bond resctricted) | 局域电荷（只关心成键原子） |
| PCWTE1 | partial charge weighted topological electronic index | TE1/qnmax |
| PCWTE2 | partial charge weighted topological electronic index (bond resctricted) | TE2/qnmax |
| LDI | local dipole index | 成键原子对电荷差和 |

1. Molecular properties (20)

|  |  |  |
| --- | --- | --- |
| Name | Description | 备注 |
| **Uc** | unsaturation count | 不饱和键的计数： |
| **Ui** | unsaturation index | 不饱和指标： |
| **Hy** | hydrophilic factor | 亲水因子： |
| **AMR** | Ghose-Crippen molar refractivity | 摩尔折射率 |
| **TPSA(NO)** | topological polar surface area using N,O polar contributions | 拓扑极化表面面积 |
| **TPSA(Tot)** | topological polar surface area using N,O,S,P polar contributions | 拓扑极化表面面积 |
| **MLOGP** | Moriguchi octanol-water partition coeff. (logP) | 正辛醇与水两相的分配系数（基于Moriguchi的回归模型） |
| **MLOGP2** | squared Moriguchi octanol-water partition coeff. (logP^2) | 平方 |
| **ALOGP** | Ghose-Crippen octanol-water partition coeff. (logP) | 正辛醇与水两相的分配系数（基于Ghose-Crippen的回归模型） |
| **ALOGP2** | squared Ghose-Crippen octanol-water partition coeff. (logP^2) | 平方 |
| **SAtot** | total surface area from P\_VSA-like descriptors | 基于P\_VSA-like描述符计算得到的总表面 |
| **SAacc** | surface area of acceptor atoms from P\_VSA-like descriptors | 基于P\_VSA-like描述符计算得到的受体原子的表面积 |
| **SAdon** | surface area of donor atoms from P\_VSA-like descriptors | 基于P\_VSA-like描述符计算得到的给体原子的表面积 |
| **Vx** | McGowan volume | McGowan定义的体积计算方法 |
| **VvdwMG** | van der Waals volume from McGowan volume | Vx基础上计算得到的范德华体积 |
| **VvdwZAZ** | van der Waals volume from Zhao-Abraham-Zissimos equation | 从Zhao-Abraham-Zissimos方程计算得到的范德华体积 |
| **PDI** | packing density index | Vx/SAtot |
| **BLTF96** | Verhaar Fish base-line toxicity from MLOGP (mmol/l) | -0.85 \* MLogP – 1.39 |
| **BLTD48** | Verhaar Daphnia base-line toxicity from MLOGP (mmol/l) | -0.95 \* MLogP – 1.32 |
| **BLTA96** | Verhaar Algae base-line toxicity from MLOGP (mmol/l) | -1.00 \* MLogP – 1.23 |

1. Drug-like indices (28)

药物评分类描述符

1. CATS 3D (300)

三维结构中药物位点的描述符。

1. 关于Intrinsic State

其中，是主量子数，是（i原子的价电子数-连H的个数），是（i原子的sigma电子数-连H的个数）（等价于去H结构的i原子的邻接矩阵的列和，也相当于去H结构的顶点指数）。