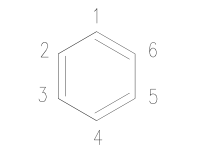
sol 转 smiles 的自动转换算法

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 编号 | SOL | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| SOL | A6 | A4 | A2 | N6 | N5 | N4 | **N3** | **N2** | **N1** | **R** | **br** | **me** | **H** | **AA** | **S** | **RS** | **AN** | **NN** | **RN** | **O** | **RO** | **O=** | **Ni** | **V** |
|  | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 编号 | SOL | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| SOL | A6 | A4 | A2 | N6 | N5 | N4 | **N3** | **N2** | **N1** | **R** | **br** | **me** | **H** | **AA** | **S** | **RS** | **AN** | **NN** | **RN** | **O** | **RO** | **O=** | **Ni** | **V** |
| 顶点 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 边 | 9 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

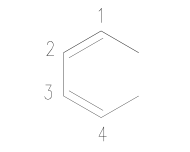
邻接矩阵

针对A6，无氢邻接矩阵AJ(A6)可表示如下

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 0 | 2 | 0 | 0 | 0 | 1 |
| 2 | 2 | 0 | 1 | 0 | 0 | 0 |
| 3 | 0 | 1 | 0 | 2 | 0 | 0 |
| 4 | 0 | 0 | 2 | 0 | 1 | 0 |
| 5 | 0 | 0 | 0 | 1 | 0 | 2 |
| 6 | 1 | 0 | 0 | 0 | 2 | 0 |

每个顶点的c-c 键数为3（C-H键忽略，每个顶点所连接的边数为4-1=3），同理列符合相同的规律。

|  |  |
| --- | --- |
| A6 |  |
|  | A4 |

若增加A4结构单元，

则新增顶点，新增c-c键（边）

A6 + A4 拼接成完整分子，需要完成两步

第一步，组成分块对角矩阵

第二步，连接A4的边与A6的顶点，即完成上对角矩阵的填充

第二步又可以分为两小步

2.1 确定主结构上的顶点，如接A4，则主结构上的两个顶点需要相邻，并有边，即

为所选的顶点

例如

可以选择、、、、、。如这里选择

2.2 连接主结构上的顶点与副结构上的开放边

开放边

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 |
| 1 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 2 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 1 | 0 | 2 | 1 | 0 | 0 | 0 |
| 6 | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 1 |
| 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 1 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 2 |
| 4 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 0 |

镜像、交换、旋转等变换矩阵

A6 邻接矩阵特征多项式

特征根为

-3.000, -1.732, -1.732, 1.732, 1.732, 3.000

图 着色

双键 单键

环数（cycle）

偶图 （dualist）异构体枚举 （isome enumeration）

异构数目规则（isomer number regularity）

dual

inner dual

planar ployhex

距离矩阵

最短距离（不同环上的顶点）

On the concept of the weighted spanning tree of dualist

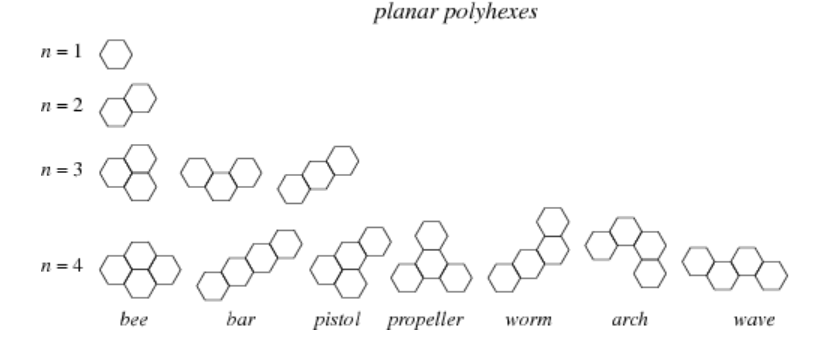
平面多六边形 planar polyhexe

n = 1 苯环

n= 2 萘

n =3 菲

n =4



bee

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| A6 | A4 | A2 | N6 | N5 | N4 | **N3** | **N2** | **N1** | **R** | **br** | **me** | **H** | **AA** | **S** | **RS** | **AN** | **NN** | **RN** | **O** | **RO** | **O=** | **Ni** | **V** |
| 1 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

wave

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| A6 | A4 | A2 | N6 | N5 | N4 | **N3** | **N2** | **N1** | **R** | **br** | **me** | **H** | **AA** | **S** | **RS** | **AN** | **NN** | **RN** | **O** | **RO** | **O=** | **Ni** | **V** |
| 1 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

在SOL的表达规则中，多环分子是以“背缩”规定生成，例三环优先生成菲，四环优先生成䓛。

稠环分子sol 表达式A6 A4 A2 的数量与偶图的关系

例一

|  |  |  |
| --- | --- | --- |
| A6 | A4 | A2 |
| 1 | 0 | 0 |

|  |  |
| --- | --- |
| 分子结构图 | 偶图 |
|  |  |

数量关系

|  |  |  |
| --- | --- | --- |
| sol数量和 | 偶图顶点 | 偶图边 |
| 1+0+0 | 1 | 0 |

例二

|  |  |  |
| --- | --- | --- |
| A6 | A4 | A2 |
| 1 | 1 | 0 |

|  |  |
| --- | --- |
| 分子结构图 | 偶图 |
|  |  |

数量关系

|  |  |  |
| --- | --- | --- |
| sol数量和 | 偶图顶点 | 偶图边 |
| 1+1+0 | 2 | 1 |

例三

|  |  |  |
| --- | --- | --- |
| A6 | A4 | A2 |
| 1 | 2 | 0 |

|  |  |
| --- | --- |
| 分子结构图 | 偶图 |
|  |  |

数量关系

|  |  |  |
| --- | --- | --- |
| sol数量和 | 偶图顶点 | 偶图边 |
| 1+2+0 | 3 | 2 |

例四

|  |  |  |
| --- | --- | --- |
| A6 | A4 | A2 |
| 1 | 2 | 1 |

|  |  |
| --- | --- |
| 分子结构图 | 偶图 |
|  |  |

数量关系

|  |  |  |
| --- | --- | --- |
| sol数量和 | 偶图顶点 | 偶图边 |
| 1+2+1 | 4 | 5 = 1+2-1+1\*3 |

关系归纳：

偶图的顶点数

偶图的边数

偶图的三角环数

验证

例四 晕苯

|  |  |  |
| --- | --- | --- |
| A6 | A4 | A2 |
| 1 | 3 | 3 |

|  |  |
| --- | --- |
| 分子结构图 | 偶图 |
|  |  |

数量关系

|  |  |  |  |
| --- | --- | --- | --- |
| sol数量和 | 偶图顶点 | 偶图边 | 偶图三角环 |
| 1+3+3 | 7 | 12 = 1+3-1+3\*3 | 6 = 2\*3 |

规律符合

2. 那么找到了这个规律后，怎么用，才能得到smiles

第一步：从偶图得到邻接矩阵

第二步：

判断存在内点与否

为全图（区别于偶图）的内环顶点数

内环点个数 = 三角形个数 =

偶图三角形的个数 = 三个苯环的公共顶点

偶图的边 = 两个邻接苯环的公共边

以偶图的边为对称轴，可以找到邻接苯环公共边的两个公共顶点，一个三角形除去三个苯环的公共顶点，另外有三个两两相互邻接的苯环公共边的两个公共顶点（有点绕）；两个三角形呢，有两个三环公共顶点和4个邻接公共顶点。

A4的个数与A2的个数有关系吗？

A2最多等于3吗？

若没有A2，即没有三角形，则偶图的边就是

1) 背缩规则(不存在A2)

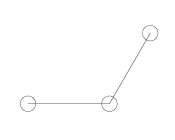
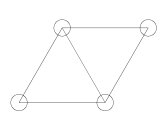
为树形结构

2) 三元环 (存在A2)

三元环结构

难点：存在异构图

若可以增加A2，则必须存在如下结构，使得增加A2 后，能组成菱形结构。

每一个三角形内部存在一个三环公共顶点，且每一个A2，均有两个共边的三角形。两个公共顶点的连接边垂直于该公共边。且是内边。

寄其中一个公共顶点为6，另一个公共顶点为7，边为（6，7）

公共顶点属于第一个六元环内，

第一步：由A6 A4 A2 组成分块对角阵

第二步：判断是否存在内点（内环），即是否存在A2>=1

若存在，if A2==1？则，内点数为2，即存在两个共边三角形，组成菱形。

第三步： 补充公共点的连接

先采用 He Matrix 生成稠环芳烃的结构

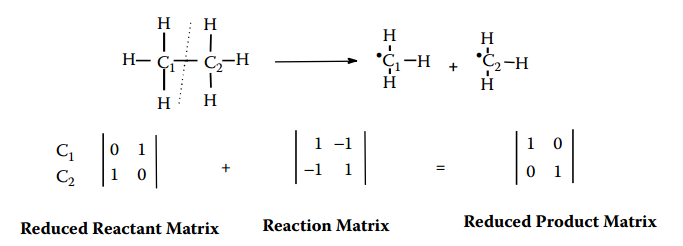
参考

何文辰《平面稠环芳烃的产生和计数》

He Matrix

线性变换 旋转 对称 平移

再借鉴键电矩阵反应矩阵reaction matrix，合并是反应的逆过程

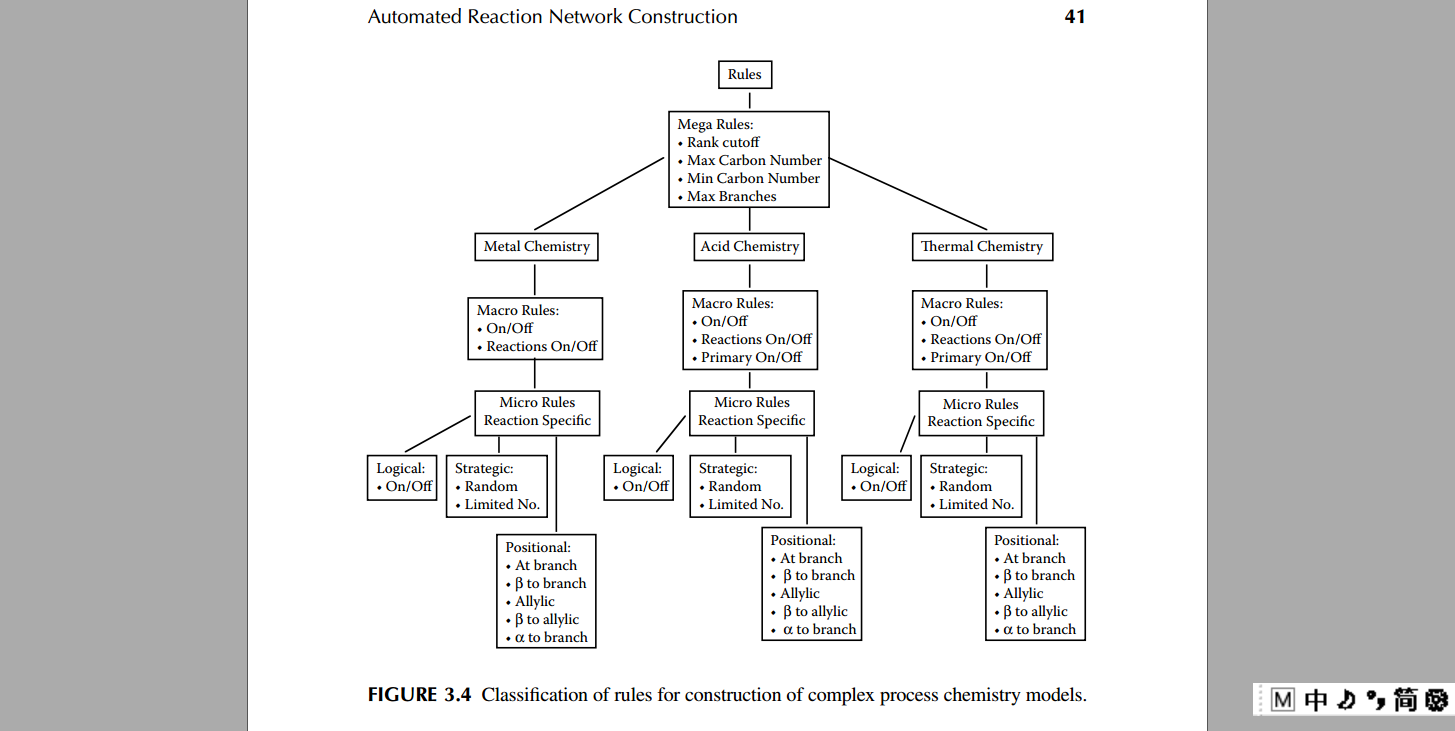


Reduced Product Matrix – Reaction Matrix = Reduced Reactant Matrix

Reduced Product Matrix 及 Reduced Reactant Matrix 均为碳骨架邻接矩阵

关键是怎么构建Reaction Matrix

参考反应网络构建规则，制定sol连接规则



3) 冠状结构

例如晕苯