

简单有机化合物同分异构体数目的计算与结构式展示

曹博涵(物)、蒙洋、江柔、何晴雨(化)

摘要

对有机物的研究十分重要但有机物数量繁多。为利用计算机快速找寻和计算有机化合物同分异构体，减少脑力计算，我们小组利用 C++ 和 Python 相关库实现了对烷烃、单烯(炔)烃、醚、单环烷烃、单取代烃(卤代和醇)同分异构的计算，及单烯(炔)烃、醚、单取代烃(卤代)的列举，并可以根据输入分子式自动分类并将所有可能的结构式展示为图片。

问题求解思路

1. 数目计算

- **单醇、卤代烃** 其数目就是相应烷基的数目，而烷基的数目可以由低级烷基的数目递归而来。
- **单醚、烯、炔** 这些化合物都具有两个烃基夹着一个键的结构。区别仅仅在于烃基顶点最大度数分别为三、二、一。利用烷基计算的结论组合即可。
- **烷烃** 根据图论，烷烃同分异构数 D_n 与标碳烷烃 B_n 、标键烷烃 C_n 、烷基数 A_n 有关系：
$$D_n = B_n - C_n + A_{n/2}$$
据此可以计算烷烃数目。
- **单环烷烃** 先对环的碳个数进行枚举。对每一类环，都用碳总数减去环上碳个数，得到取代基的碳总数，然后利用烷基数目的结论，通过数组枚举取代基碳数不同的情况，最后加和起来。

2. 列举

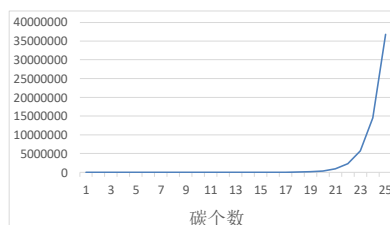
- **单取代烃、醚和烯炔** 仿照数目计算的思路，只需将计算数目的过程改成拼接烃基(本质上是树的连接)的过程。但烷烃的减法无法实现，得另辟蹊径。
- **烷烃** 烷烃是主链及其上的烷基，而烷基是一棵树，最长链长度即为深度。若深度大于距离端基碳的距离，则改变主链，必重复；小于必不重复；若等于则需考虑较多因素，因而最后并未完全实现。但前两种情况已经得到了考虑。
- **炔烃** 把炔烃的 C#C 当成整体，可以将炔烃看作醇和醚的混合，因而不必增加函数来实现，只需调用原有函数，并进行转换。

3. 列举的展示

- **Python/C API 调用** 为提高效率，我们使用 C++ 编写数目计算和列举算法，并将列举结果用 SMILES 字符串表示。再用 Python/C API 将其转化成 Python 可调用的函数库。
- **自动分类** 对用户输入的分子式，计算不饱和度及判断杂原子后对其分类，再调用相关函数计算展示。
- **tikinter** 我们使用 Python 自带的图形界面库 tikinter，及其中 Entry 等组件实现与用户的交互。
- **RDKit** 将列举结果由字符串转化为图片，实现可视化展示。

问题求解效果

• 数目计算结果



经过计算，我们得出了 25 个以下的各类有机物数目(大于时溢出)，并做出数据如图。对于烷烃，我们发现其趋势是指数增长，这是有机物数量繁多的原因之一。

• 图形界面列举效果



左图是最终的图形界面效果，可以根据需要选择性输出数目、结构式(SMILES)、结构示意图。右图是用户输入 C5H11Cl 后的结果。

遇到的困难与解决

1. 我们大量学习了有关 Python 的知识；
2. 我们测试了有关递归与循环算法的效率，并在适当的地方使用了循环算法；
3. 利用类的思想将烷基当做一个类，其方法有拼接等，从而实现了简化；
4. 利用 Gitee 代码仓库和多文件方法实现了多人协助编程及版本管理；

待解决问题思路

1. 利用大数计算和 SQL 实现更大数目计算及加速；
2. 利用其他算法实现多种计算并考虑立体异构。

参考资料

- [1] Python 上的化学信息学开源工具包 RDKit 说明文档 [DB/OL]. <http://www.rdkit.org/docs/index.html>
- [2] 知乎网友走地鸡. 烷烃同分异构体数目的初等叙述方式 [Z]. <https://www.zhihu.com/question/264476342/answer/1075108614>
- [3] 伍启期. 一类环烷烃的异构体个数的计算公式[N]. 佛山大学佛山师专学报(理工版). 1990,(02)