化学学报 2001 59 (2): 241-246 ISSN: 0567-7351 CN: 31-1320/06

目标化合物析分系统中的重要算法

王利莎,袁身刚,欧阳政,郑崇直

中国科学院上海有机化学研究所.上海(200032);中国科学院计算机化学开放实 验室.上海(200032)

收稿日期 修回日期 网络版发布日期 接受日期

摘要 介绍了目标化合物析分系统中所用到的三个重要算法。它们是：最短拓扑距离的求解、
析分过程结束的判别以及合成树的构建。分子结构中任意两个原子之间最短拓扑距离的求解是建立在采用队列数据结构的宽度优先搜索算法基础上的，
析分过程结束的判别是由在新Morgan算法基础上产生的化合物的唯一编码和B－树两种算法构成，
前者是为了将前体与原料库中每个化合物是否同构的复杂问题简化为码之间的比较问题；后者是一种高效文件组织方式，
将代表原料库中化合物的唯一编码作为检索键来组织建库，从而实现对原料库的快速查询，合成树采用的是链表存储方式，每一个结点由六个域组成，
且在建树和画树的过程中，均用前序遍历。这些算法是实现析分系统的基础，因此它们的正确设计与高效实现就显得尤为重要。

关键词 目标化合物 计算化学 算法 计算机化学 拓扑距离

分类号 O-39

Important algorithms used in the target parsing system

Wang Lisha,Yuan Shengang,Ou Yangzheng,Zheng Chongzhi

Shanghai Inst Organ Chem., CAS.Shanghai(200032);Lab of Comp Chem, CAS.Shanghai(200032)

Abstract This article presents the three most important algorithms used in the Target Parsing System. They are the shortest topological distance between two atoms, the determination of the termination of the parsing process and the construction of the synthesis tree. The shortest topological distance between two atoms in a structure was designed according to the widely used breadth first search algorithm which uses a data structure called Queue. The determination of the termination of the parsing process consists of the canonicalisation algorithm based on the new Morgan algorithm and the B - tree algorithm, The former is used to convert the complicated isomorphic problem into the comparison between two unique codes. The latter is one of efficient file organizations, in which the unique code is used to represent the compound in the material data base. All these codes are organized in a B - tree. Thus, for a given compound it is very fast to know if it is contained in the data base by searching in the B - tree. The construction of the synthesis tree adopts data structure of list. Each node of the list consists of six fields. Both the generation and drawing of the synthesis tree use the preorder technique. Since these algorithms constitute the core of the Target Parsing System, their proper design and implementation is extremely important.

Key words COMPUTATIONAL CHEMISTRY ALGORITHM

DOI:

通讯作者