

计算机辅助质谱预测和候选结构评价

洪群发 朱大模 杨伯宇 许崇德 卢保章

(中国科学院大连化学物理研究所)

[摘要] 本文介绍一个用基本的质谱裂解规则来预测已知结构化合物的质谱的程序。如果给定对应某未知物的一组候选分子结构,该程序可通过预测质谱的比较来评价这些候选结构。程序采用动态描述方法存储裂解规则,可自动进行各种结构类型化合物的质谱预测。

关键词: 质谱预测 候选结构评价 裂解规则

一个完整的智能结构解析系统,其目的就是要向分析者提供有未知物的正确的分子结构信息。但经由结构产生程序产生的候选分子结构一般不止一两个,对于大的复杂分子往往会产生大量的候选分子结构。在这种情形下,正确地预测各候选分子结构的谱图特征,通过比较预测谱图和实验谱图的相似程度进行候选结构评价,对减少系统最终输出结果的数目,正确地鉴定未知物的分子结构,是一个极有价值的方法。DENDRAL 系统^[1,2]、CASE 系统^[3]和 CHEMICS 系统^[4]都发展了自己的谱图预测程序。

一般谱图预测程序都采用“亚结构→谱图特征”规则来预测候选结构的谱图特征,这种方法受到“亚结构→谱图特征”规则的数量和质量的限制。DENDRAL 系统开始采用了一种称为“半级理论”^[5]的模式来预测质谱。但“半级理论”中用来预测质谱的键断裂过程描述过于一般化,而且该方法只有在有经验的专家的指引下,才能发挥有效的作用。后来,他们又发展了一个基于特定类型的质谱裂解规则的质谱预测程序^[2]。但对详尽的特定类型的裂解规则的需要限制了该方法的实际应用。

本文通过对各种结构类型的化合物的质谱裂解行为的研究,采用了一种新的裂解规则描述方法,发展了一个可自动进行各类化合物质谱预测的程序。

方 法 原 理

1. 质谱预测程序

本程序是通过对质谱实际形成过程的模拟来进行质谱预测的。它采用程序化的质谱裂解规则模拟开裂化合物的分子结构,然后通过计算裂片离子的质量和裂解过程中的氢

1991年6月26日收

迁移数及中性丢失质量数来预测化合物的质谱。程序主要遵循以下两个原则:

- (1) 程序要尽量广泛地适用于各种结构类型的化合物的质谱预测。
- (2) 程序要考虑到相同的结构特征在不同的结构环境下因裂解过程的氢重排情形的不同而导致的谱图特征的差异。

基于以上原则,本程序采用了动态的质谱裂解规则描述方法。一条质谱裂解规则的一般表示形式为:

结构模型——裂解过程描述

其中,结构模型定义了断裂键的结构环境。裂解过程描述包括对结构模型中的断裂键所在位置的标识;裂解后电荷所在位置的隐性描述;以及裂解过程中伴随的氢迁移数和中性丢失质量数。



图 1 质谱裂解规则示例

图 1 给出了一条关于脂肪醇的 β -键裂解规则。图 1 中, BREAK 定义了 C1 和 C2 之间的键为断裂键,裂解后正电荷保留在 C2 上。PEAKGROUPS 定义了伴随该裂解过程可以没有氢迁移(中性丢失质量数为 0),也可以丢失一个 $-OH$ (中性丢失质量数为 17),还可能丢失一个 H_2O ($H+OH$, 中性丢失质量数为 18)。

为了加强规则的通用性,本文采用了动态规则描述方式来描述裂解规则。一方面用动态符号来表示规则中支配裂解过程的原子(或原子团)。如图 1 中的裂解规则,将其支配裂解过程的基团 OH 用动态符号 t 来表示,就可以把该规则的适应范围扩大到所有的 β -单键裂解过程。另一方面采用不确定的氢迁移和中性丢失描述。如图 1 中的规则就给出了 3 种可能的氢迁移和中性丢失情形。目前,程序初步采用了各种结构环境下共 77 条基本的质谱裂解规则来进行质谱预测。

2. 候选结构评价程序

候选结构评价是通过比较各候选结构的预测质谱和相应的实验质谱的相似程度来进行的。相似度 Si 的计算公式为:

$$Si = 1000 - (SUMA + SUMI)/n \quad (1)$$

式中, SUMA 是实验谱图中没有匹配上的峰的强度权值之和, SUMI 是预测谱图中没有匹配上的峰的质量数权值之和, n 是两张谱图中匹配上的峰的数目。

由于质谱裂解规则的描述中, PEAKGROUPS 采用了不确定描述来定义伴随裂解过程的氢迁移数和中性丢失质量数。所以,应用一条裂解规则,程序将产生一组(而不是一个)相邻的质谱峰。所以,候选结构评价程序在考虑预测质谱对相似度 Si 的贡献时,认为在一组预测质谱峰中,只要有一个峰和实验谱图匹配成功,即认为该组预测质谱峰和实验谱图匹配成功,这组峰内其它没有匹配上的峰就不计算在(1)式中的 SUMI 内。如果一组预测质谱峰都没有匹配上,其 SUMI(i) 为:

$$SUMI(i) = [FMZ(i) \times LMZ(i)]/MW \quad (2)$$

式中 FMZ(i) 是该组第一个峰的质量数, LMZ(i) 是该组最后一个峰的质量数, MW 是对应候选分子结构的分子量。

各候选分子结构的预测质谱和实验质谱的匹配相似度将用来对候选分子结构进行排序, 也可用来帮助鉴别各候选结构的相对可靠性。

程 序 实 现

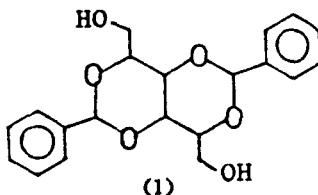
本程序主要包括以下几个部分:

(1) 候选分子结构的结构类型识别: 即对支配某种质谱裂解方式的结构特征的识别。这些结构特征(如季碳、叔碳、不饱和碳、芳香环、杂原子、甾环、杂环等)按能被计算机识别的形式存储在程序中, 程序用这些预定义的结构特征对候选分子结构进行相容性匹配, 以确定候选结构中是否存在哪些结构特征, 并产生与这些结构特征相对应的质谱裂解规则集合。一个分子结构中, 往往含有不止一种结构特征。

(2) 基于裂解规则的质谱预测: 首先程序把规则中的结构模型和候选分子结构进行匹配, 匹配成功之后按规则中 BREAK 定义的裂解方式产生相应的裂片离子, 然后按 PEAK-GROUPS 所定义的不同氢迁移数和中性丢失质量数产生一组特征质量数。所有的裂解规则对一候选分子结构的处理结果, 加上根据该结构的大小和稳定性等因素得到的关于分子离子峰的有无, 就形成了该候选分子结构的预测质谱。

(3) 候选结构评价: 比较各候选结构的预测质谱和未知物的实验质谱, 并按(1)式计算两谱图的相似度 Si, 最后按各候选结构的 Si 值的大小顺序排队、输出。

结 果 和 讨 论



1. 程序的质谱预测能力

在此我们通过比较化合物 Glucitol, 2,4:3,5-di-O-benzylidene-, D-(1) 的标准质谱和程序对该化合物的预测质谱来验证本程序的质谱预测能力。图 2 给出了该化合物的质谱图。对该化合物的质谱预测结果列于表 1。

表 1 化合物 Glucitol, 2,4:3,5-di-O-benzylidene-, D-的质谱预测结果

30	31	59	60	61	76	77	78	105	106	107	121	122	123	149	150
151	191	192	193	207	208	209	280	281	282	325	327	328	340	341	342
357	358	359													

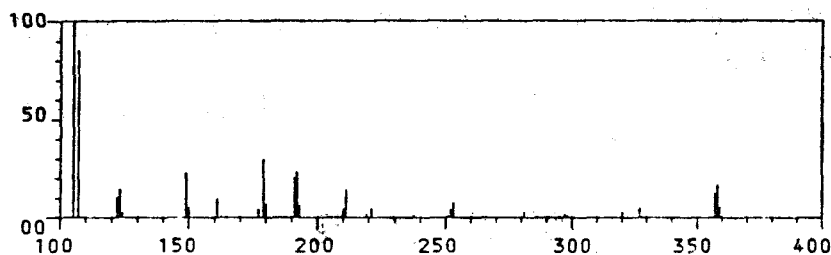


图 2 化合物 Glucitol, 2,4:3,5-di-O-benzylidene-, D-的质谱

对照分析该化合物的质谱裂解过程和质谱预测程序预测结果可以看出,对该化合物,除了 M/Z 179 是经历双环三键裂解形成的而没有被程序考虑外,其他裂解过程所形成的质谱峰都正确地预测出来了。由于某些原因该化合物的质谱没有保留 M/Z 小于 100 的峰,所以预测质谱中质量数小于 100 的峰在实验谱图中没有出现。另外,在质谱预测过程中,程序目前还没有考虑到预测质谱峰的峰强度。

2. 质谱预测和候选结构评价

为了进一步验证本程序的质谱预测能力和结构特征的识别能力,我们用本程序处理了 Mono-oxoandrostane (2) 的所有 11 个异构体。做法是任选一个异构体作为未知物,把包括该异构体在内的 11 个异构体作为该未知物的候选分子结构,然后进行质谱预测和候选结构评价,其结果列于表 2。

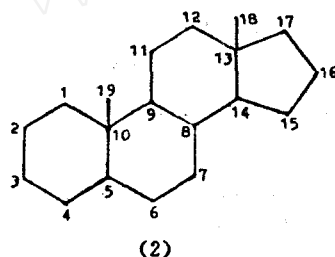


表 2 程序对 11 个 mono-oxoandrostane 异构体的评价结果

结 构	谱 图										
	1-oxo	2-oxo	3-oxo	4-oxo	6-oxo	7-oxo	11-oxo	12-oxo	15-oxo	16-oxo	17-oxo
1-oxo	637	504	755	562	480	703	874	601	923	856	668
2-oxo	637	504	755	562	480	703	874	601	923	856	668
3-oxo	637	504	755	562	480	703	874	601	923	856	668
4-oxo	637	504	755	562	480	703	874	601	923	856	668
6-oxo	458	356	607	322	534	699	888	500	936	769	524
7-oxo	384	259	555	435	306	713	832	659	933	819	671
11-oxo	482	355	606	410	405	669	888	589	937	809	635
12-oxo	376	249	551	422	413	588	757	615	943	760	644
15-oxo	414	330	581	494	292	553	793	515	967	789	674
16-oxo	362	218	432	411	209	505	824	570	936	793	668
17-oxo	414	330	581	494	292	553	793	515	967	789	674

表 2 中,12-oxo 异构体的正确结果排在第二位,16-oxo 异构体的排在第六位,其他的

异构体的正确结果也都以第一或并列第一的形式输出。16-oxoandrostane 由于其预测质谱的峰数少,虽然质谱预测结果的所有峰在各参考谱图中均能很好地匹配,但得分(即相似度)相对来说还是较低。在和所有 11 张参考谱图的比较中,16-oxo 异构体的得分都排得比较靠后。

DENDRAL 系统的质谱预测程序也曾做过同样的试验^[1],它对 2-oxoandrostane 没有给出可靠的结果,在其他 10 个异构体中,除 17-oxandrostane 外,其余的异构体的正确结果均以第一或并列第一的形式输出。

从两个程序的质谱预测和候选结构排序的结果比较可以看出,本程序和 DENDRAL 程序的质谱预测能力相当,但 DENDRAL 程序是在专家的参与下进行的(要由专家通过亚结构模型来定义被处理的分子结构中哪些裂解过程发生的可能性较大);而且,每定义一次参数,只能处理一种(或相似种类的)结构类型的化合物。而本程序是用基本的预定义裂解规则来自动进行质谱预测的(如对 Mono-oxoandrostanes,程序就采用了裂解规则集中有关饱和脂肪环的那部分裂解规则)。可预测各种结构类型的化合物的质谱。所以,本质谱预测程序具有更广泛的实际应用能力。

参 考 文 献

- [1] N. A. B. Gray et al. , Anal. Chem. , 52, 1095(1980)
- [2] A. Lavanchy et al. , Org. Mass Spectrom. , 15, 355(1980)
- [3] C. A. Shelly et al. , Anal. Chem. , 50, 1522(1978)
- [4] I. Fujiwara et al. , Anal. Chim. Acta, 133, 527(1981)
- [5] D. H. Smith et al. , Tetrahydron, 29, 3117(1973)

Computer Aided Mass Spectrum Prediction and Structure Evaluation

Hong Qunfa, Zhu Damo, Yang Boyu, Xu Chongde, Lu Peizhang

(Dalian Institute of Chemical Physics, Academia Sinica, Dalian 116012, PRC)

Received 26, June 1991

Abstract

A program is described for predicting the mass spectrum of a given molecular structure based on basic fragmentation rules. Given a set of candidate structures for an unknown compound, this program can be used to evaluate each candidate based on a comparison of the predicted and observed spectra. A dynamic method is used to describe the fragmentation rules, and the program can predict mass spectra automatically for compounds of various kinds of structures.

Keywords: Mass spectrum prediction, Structure evaluation, Fragmentation rules.