

质谱解析系统 ASES/MS 的新发展

朱大模 杨伯宇 洪群发 孙国强 许崇德 刘人余

(中国科学院大连化学物理研究所)

一、引言

ASES/MS 系统^[1]自问世以来又有了重大发展,这主要是向系统引入了学习功能。初始的 ASES/MS 系统使用专家提供正离子的质谱规律和中性丢失规律两个知识库中的质谱规律解释质谱,推导分子结构,系统具有推理功能。由于系统没有学习功能,也就没有自身生产质谱规则的能力。所以尽管有两个庞大的知识库,使用起来还是感到结构信息不足。学习功能的引入,使得计算机就象人类专家一样,能够带着未知质谱去从数据库查找出相似的质谱图,进行学习并总结出亚结构/质谱联系规则,供阐明分子结构用。这样,系统甩掉了两个庞大的知识库,用学习数据库质谱方法直接应用 45,000 个各类化合物所含的丰富结构信息于结构鉴定中。其次,是向系统引入了功能团质谱法技术,系统能自动地从试样质谱产生和应用一级至 N 级中性丢失质谱,从中获得分子内相对电负性较强的那部分亚结构的结构信息。经常这类结构信息是不可能直接从正离子质谱获得的。第三,随着结构信息提取方式的完善,我们又完善了分子结构式产生程序,新的结构产生程序,对亚结构的化学价没有限制。它能依据系统推得的亚结构详尽地、不重复地产生符合预测分子量的全分子结构异构体。计算机学习功能和功能团质谱法的引入是智能系统的根本标志。ASES/MS 系统开发利用 EI 质谱所固有的丰富结构信息,达到了新的高度。

二、系统简介

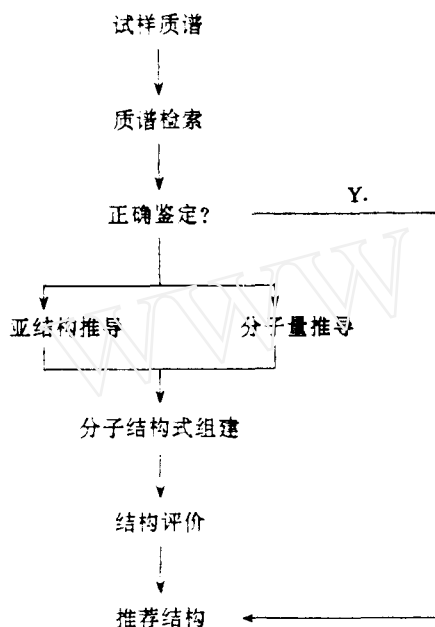
ASES/MS 系统(图表 1)串联检索系统和解释系统鉴定试样物质的分子结构。检索系统采用独立发展的正反检结合型原理,鉴定试样物质为质谱数据库内 45000 个化合物之一。如果,检索鉴定失败(库中无此化合物),解释系统应用质谱理论,从样品质谱出发,推导分子量;学习数据库质谱,由此总结出亚结构/质谱联系规则,并用以推导试样分子中可能含有的亚结构;分子结构组建程序,按一切可能方式,从这些推得的亚结构,不遗漏也不重复地产生试样分子的全部结构异构体。对每一异构体预测质谱,并将它们与试样的实验质谱比较做结构评价。最后,按优劣顺序输出候选结构。

化合物分子的任何一部分结构都叫亚结构,它可以是有机化学功能团(包括基元结构,如 $-\overset{|}{\text{CH}}-$, $-\text{CH}_2-$...)及其组合形式。所以分子结构 $\geq$ 亚结构 $\geq$ 功能团、基元结构。正确的数据库质谱,是用检索方法正确鉴定样品物质的基础。不遗漏地鉴定出样品分子所含的全部亚结构是解释系统达到正确结构鉴定的关键一步,数据库质谱(教科书)的正确性、所含结构类型的完备程度、及对计算机学习功能的培养程度是完成这关键一步的关键。

三、系统的新发展

用于开发利用已经积累起来的大量质谱数据所含的丰富结构信息的方式及所能达到的程度,是对计算机质谱法结构鉴定系统进行科学分类(检索、模式识别和智能系统)和评价其水平的依据。令计算机自己去学习质谱,总结亚结构/质谱联系规则,并用于解释质谱,从试样质谱获取样品分子的详尽亚结构信息,是 ASES/MS 系统的主要新发展。

图表1. ASES/MS系统逻辑



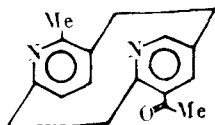
1. 质谱信息的机内表示

ASES/MS 系统的质谱数据库含 45,000 化合物的质谱数据,每一化合物的库存信息是:化合物名称、分子结构式、分子式、分子量和 EI 质谱数据。分子结构式-质谱数据是最根本的信息对。计算机能看懂质谱数据和分子结构式,是计算机学习质谱,总结规律的前提。质谱数据是数字量,十分便于计算机存储与阅读。分子结构式在 ASES/MS 系统中是平面图形。用 ASES 线性结构编码法^[2]存储于计算机中,为一字符串。系统内有一计算机程序,当需要时它自动把有关的那些结构式(ASES 码)转换为 5×N 结构邻接表(图表 2),实现计算机对分子结构式的识别和处理。N 为分子中非氢原子(ASES 码中的基本结构标记符)的数目。

图表 2 ASES/MS 中的分子结构式表示

化合物名称: 6-METHYL-3'-ACETYL-(2,2')(5,5')-PYRIDINOPHANE

结构式:



ASES 线性码: Y (M) N Y * X X Y * ; V V Y * N X Y * V V ; Y (C(O)M)X.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

5×N 邻接表:

节点→

邻接节点

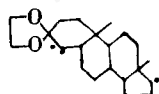
Y1	M2	N3	Y4	X5	X6	Y7	V8	V9	Y10	N11	X12	Y13	V14	V15	Y16	X17	C(O)18	M19
M2	Y1	Y1	N3	Y4	X5	X6	Y4	V8	V9	Y10	N11	X12	Y13	V14	Y10	Y16	Y16	C(O)18
Y7		Y4	V8	X6	Y7	Y1	V9	Y10	N11	X12	Y13	V14	V15	Y7	X17	Y13	M19	
N3			X5			V15		Y16			X17				C(O)18			

2. 模式识别亚结构/质谱联系规则产生系统

在缺乏分析样品的背景材料时,质谱分析者从实验谱图出发,通过分子量、质谱基峰、“特征离子”查询等,便可从质谱数据库查找出一些同实验质谱十分相似的“标准”质谱,使问题获得解决或者找到解决问题的线索。在这后一种情况,分析者总是假定:同样品质谱有共同质谱峰(或者再考虑到取代基的质量位移影响)的标准质谱化合物分子,与样品分子之间有某种共同结构存在。用模式识别技术产生亚结构/质谱联系规则,就是要计算机模拟人类专家求解问题的这个本领。

把化合物质谱作分段比较,利于鉴定分子亚结构,离子系列质谱与分子类别特征有更多联系。综合这两方面的优点,系统采用分段式离子系列质谱来表示每一化合物的质谱信息,建立标准模式集。应用模式识别的 KNN 技术实现把同样品质谱最相似的数据库数据(分子结构/质谱数据对)集合起来,用人工智能的启发式搜索方法,计算机就能由比较两个分子结构式提取最大公共结构,最后获得由 KNN 推荐的那组分子结构式的最大公共结构,将它与最大公共结构质量数相匹配的一组共同质谱峰相结合,就得到亚结构(最大公共结构)/质谱特征联系(图表 3)。

图表 3 用模式识别法产生的亚结构/质谱联系规则示例
亚结构:OVVQC;QT*Q*(M)VV;VVT*T*;T*Q*(M)VV;VVT.



质谱特征:		M/Z	W	M/Z	W	M/Z	W
	41	179	227	269	0	0	0
	43	154	231	667	0	0	0
	55	218	245	269	0	0	0
	81	372	247	282	0	0	0
	95	154	0	0	0	0	0
	99	1000	0	0	0	0	0
	100	397	0	0	0	0	0
	139	974	0	0	0	0	0
	161	346	0	0	0	0	0
	177	282	0	0	0	0	0

3. 以键断裂规则为基础的亚结构/质谱联系规则产生系统

用模式识别方法产生的亚结构/质谱联系规则,有很好的统计性,并能与亚结构所处的化学环境相联系。但是,方法的成功应用是以标准模式集中含有足够数目的、处于相似环境的同一亚结构的化合物质谱为前提的,这对相当一部分亚结构来说,还是一个苛刻的条件。质谱分析专家,以质谱理论为依据,借助于一系列键断

裂过程和重排过程关联分子结构与其质谱,推导亚结构/质谱联系规则就不受这种限制。

标准质谱分析方法中使用 50—70eV 能量的电离电子,它足以从分子内的任何化学键或含孤电子对原子击出一个电子。因此,在各分子离子之间存在着按正电荷和自由基位置的分布。诸位置的电离电位,诸位置及其环境对正电荷及自由电子的稳定能力决定着这种分布的性质。自由基位置和过剩能量的引入导致生成分子离子的不稳定性,它趋向于解裂为小的裂片离子和中性裂片。质谱数据指示:分子中的叔碳、季碳、不饱和键、含孤电子对的杂原子常常是这种正电荷和自由基位置中心,它们的 α , β 有时 γ 位置化学键有很强的断裂趋势。此外,生成五、六员环离子的特殊稳定性,尤其是生成共轭稳定的环状离子的情况,还导致在平常条件下不易断裂的键断裂,生成丰富离子,已经为系统发展了一个计

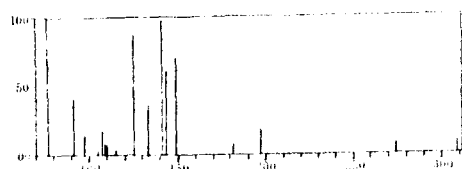
算机程序。它能识别分子结构的上述特征,处理相应的断键过程,按有氢和无氢原子迁移的断键过程关联分子结构式与质谱。最后,把与同一亚结构有关的诸离子归总起来,生成亚结构/质谱联系规则。图表 4 的结果与人类专家关联分子结构与其质谱所得的结果相符合。

图表 4 计算机模拟用断键过程解释质谱
推导亚结构/质谱联系规则示例

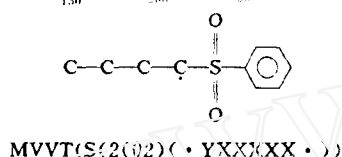
化合物名称:1,1-BIS(PHENYLSULFONYL)-BUTANE

质谱图:

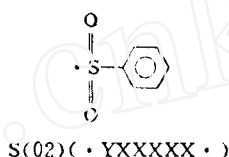
计算机推得的亚结构/质谱联系规则:



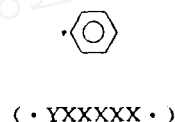
197 MVVT(S(02))(·YXXXXX·) 197 143 141 125 109 77
141 S(02)(·YXXXXX·) 143 141 125 109 77
77 (·YXXXXX·) 77



197



141



77

4. 以功能团质谱法为基础的亚结构鉴定系统

组成的分子的亚结构(自由基)都有自己特定的电离电位,按照 STEVENSON 规则,相对电离电位较低的亚结构,易生成正离子,便于直接用正离子质谱法研究其结构。反之,相对电负性较强的亚结构,难于生成较强的正离子,它们的结构信息,只能从一级至 N 级中性丢失质谱推得。

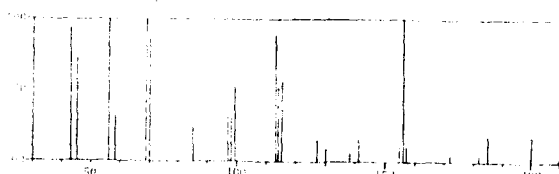
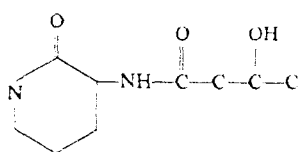
研究一级至 N 级中性丢失质谱与功能团结构的联系,我们已经推导出各种有机化学功能团以及不同化学环境中同一功能团的质谱特征-功能团质谱^[3]。定义亚结构为有机化学功能团及其组合形式,我们就能从有限种类和数目的功能团质谱出发解释试样质谱,推导出无限数目的样品分子中可能包含的无限数目亚结构形式。为此发展的计算机程序,它能自动简化样品质谱为 N 个有意义峰,计算出一级至 N 级中性丢失质谱共 N-1 张,并使之与功能团质谱相匹配,检出样品分子可能含有的功能团,然后再反过来参照 N-1 张中性丢失质谱,从这些检出的功能团中选择合适者组合成亚结构,评分给出(图表 4)。

图表 5 用功能团质谱法推导亚结构示例

化合物名称: BUTANAMIDE, 3-HYDROXY-N-(2-OXO-PIPERIDIN-3-YL)-

结构式:

质谱:



检出功能团:		组建亚结构:	
一价	二价	新建亚结构	相对可靠性
-M	-T(OH)V-	-NHC(O)CH ₂ CH(OH)CH ₃	140
-T(OH)M	-TO(OH)-	-N(M)CH ₂ NHC ₂ H ₅	137
-NM ₂	-C(O)N(H)-	-OC ₂ H ₅	136
-C(O)OH	-N(H)C(S)N(H)-	-NHC(O)C(O)CH ₂ OH	131
-VVOH		-N(M)CH ₂ N(CH ₃) ₂	130

四、小 结

计算机学习功能是智能系统的根本标志。用计算机模拟人类专家学习质谱,总结结构/质谱联系规则,直接从基础数据库质谱出发,去解释基础数据库中无标准质谱的真正未知物的实验质谱,阐明分子结构,这种“计算机学习”法比知识库方法有更强的学术性,解决问题更客观。因为它所依据的基础数据库质谱,就是纯化合物的实测质谱,在仪器型号、实验条件范围内是可重现的,不因人而异。所以,人们对基础数据很少有更改之事,新增数据则是经常的。反之,知识库知识则因专家资历而异,甚至 A 专家可能会对 B 专家提供的知识完全否定。当然,计算机的学习功能是人类专家赋予的,受专家资历的限制。但是,一旦专家对某个问题有了新的认识,他只需修改计算机程序就可能把系统全盘更新到新的高度上来。所以,解释质谱的“计算机学习”系统比知识库系统具有解决问题客观,管理费用低,十分便于更新换代的优点。

参 考 文 献

- [1] Zhudamo, Wang Luoqiu et al, ASES/MS: An Automatic Structure Elucidation System for Organic Compounds Using Mass Spectrometric Data, 《Analyst》113(1988)1216
- [2] 朱大模、余建文等, ASES 系统中的化学结构编码和数据紧缩技术, 《计算机与应用化学》6(1)(1989) 46
- [3] 朱大模、洪群发等, 功能团质谱法研究, 《分析测试通报》9(3)(1990)80