

吲哚生物碱的质谱数据检索系统

薛兰珍 黄丽瑛 徐为寒* 周韵丽** 黄知恒
(中国科学院上海药物研究)

[摘要]本文报道一个小型吲哚生物碱质谱的检索系统的编制,它包括:(1)文件库,载有由文献收集以及自行测制吲哚生物碱的低分辨质谱 1600 张,所有谱均按 biemann 原则处理,即每 14 个质量中选取 2 个最强峰,并以此组成简化谱;(2)检索程序,它可进行指纹检索与分类检索(分子量、分子式、特征碎片与特征中性丢失);(3)检索结果以数据表格或棒图方式输出,文件库可通过 DELETE 与 RENEW 指令进行修改。

一、前 言

吲哚生物碱是一类重要的生物碱,其分布广泛,结构复杂,生理功能多样,它们中的一些代表性化合物具有抗癌(如长春碱、长春新碱、长春酰胺),降压(如利血平),心血管舒张(如长春胺)的作用,并早已引用于临床。目前已知的生物碱总数达 5,000 余种,而吲哚生物碱即占三分之一以上^[1]。新的同类化合物不断出现,从中获得有苗头的药物的趋势正方兴未艾。因而,对这类生物碱的研究始终是有机化学与药物化学的一个热点。

质谱是天然产物化学分析的有力工具,它具有样品消耗少($1\mu\text{g}\sim 1\text{ng}$)。结构信息丰富的特点。特别对于生物碱类化合物,其图谱-结构关系的专一程度(SPECIFICITY)很高,由一张低分辨谱可以得出有关分子量、分子式、基本骨架以及部分结构的信息。值得注意的是,由此作出结构鉴定大多在极微($\mu\text{g}\sim \text{ng}$)数量的样品下完成的。

二、SMDI 系统的简介

SMDI 具有以下特点:

- (1)尽可能完备的数据文件:我们总集了 1960—1985 年发表的图谱约 1600 张(有 10%为我们实验室测试结果),与同类的手册^[2,3]相比,它有最大的复盖面;
- (2)灵活多样的检索方式:可以根据分子量全谱作指纹检查,也可以分别对低质量区、

1988 年 8 月 30 日收

- 中国科学院上海分院计算中心
- 通讯联系人

高质量区离子以及中性碎片作分类检查,因而可能进行全结构鉴定,或在不包含在数据文件中的未知物的部分结构作出推测;

(3)简易直观的操作方法:系统提供“菜单式”和“指令直接输入式”两种操作方式,分别导入相应的功能模块。

三、SMDI 系统的结构和主要功能

整个系统建立在 BURROUGHS 6900 计算机上,运行在中国科学院上海统一计算机网络-IDC 环境中。它为联网的各研究所共享。所有的联网用户均可在各自的实验室中通过电话线借助于智能工作站(如 PDP-11, IBM-PC, B20 等)或网络终端直接查询本数据库。其结构示如下图 1。

两个主文件的概貌分述如下:

INDOLE-MS DATA 文件

文件储有 1600 张低分辨质谱图,按 BIEMANN 标准简化,即用人工选择,每 14amu 范围内录入 2 个最强峰。数据以统一格式存放:

(1)图谱的描述信息,包括谱号、分子量、分子式、命名,均以定长格式存放在谱的第一行记录中。

(2)图谱的数据信息:质量(m/z)-相对峰强(%)的数据对,存放在每谱的第二到第十九行记录中。

(3)每谱的总峰数存放在第二十行记录中。

(4)数据来源:参考文献用 CAS 规定的略写符号书写,存放在谱的第二十一行中。

SEARCH 文件:

文件储有本系统的查询、检索功能模块程序,其主要内容为 1)主程序,2)六块检索功能块程序,3)输出方式选择功能模块程序及几个共享子程序。

(1)主程序:

考虑到本系统的特点:

1)灵活、多样的检索功能,系统程序以模块结构的方式设计,以一主程序来调度和设施各功能模块程序。各功能块程序除调用一些共享的子程序外,相互并不调用,各自完成独立功能后返回主程序,从而系统程序在结构上层次清晰,便于调试阅读。

2)简易、直观的操作方法,系统允许两种操作方式:

0: INTERACTIVE MODE (菜单选择操作方式)

1: DIRECT INPUT OF PARAMETERS(直接输入指令参数操作方式)

用户可键入“0”或“1”来选择任一种操作方式。用“0”(人机对话方式)时,操作始终在屏幕提示下进行,用户只要用“Y”“N”等简单字母及数字回答系统的询问,就能完成各项检索功能。用“1”(指令参数直接输入方式)时,用户可直接输入各指令及所规定的参数来完成相应的检索功能,其特点是直观快速。用户选定操作方式后系统即进入检索并在屏幕上显示相应的操作模式。

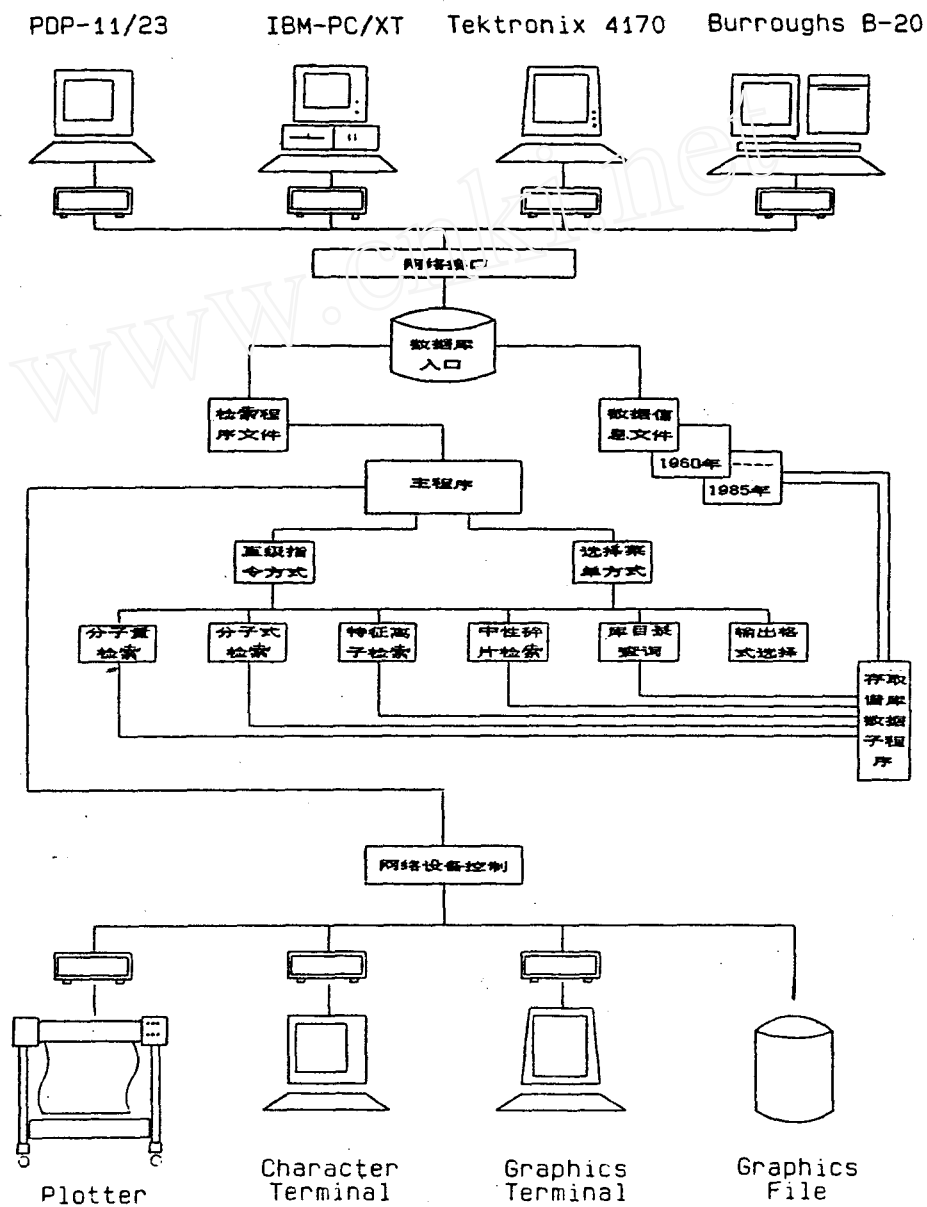


图1 系统结构图

(2)功能模块程序:

系统含二个查询功能模块(ILIST, RLIST),四个检索功能模块(MWSER, MFSER, MISER, NLSE),和二选择谱图输出格式功能模块(MLIST, BPLOT),各功能模块的主要功能和算法结构分述如下:

1) ILIST 和 RLIST——开列谱库目录表

系统按用户要求列出全部或部分目录。目录表可以按谱号由小到大排列,或按谱的分子量值,由小到大排列。ILIST 模块区别于 RLIST 模块的是,后者给出含有相应的参考文献目录,并能按化合物名称检索(NAME SEARCH),输出结果分别列于表 1、2。

表 1 ILIST 目录

LIST OF SPECTRA INDEXED BY ID #: 1-3				
TOTAL NO OF SPECTRA IN THE FILE: 1597				
ID #	MW	MF	NAMES	
1	174	C ₁₁ H ₁₄ N ₂	N-METHYLTRYPTAMINE	
2	196	C ₁₃ H ₁₂ N ₂	1-ETHYL-B-CARBOLINE	
3	198	C ₁₂ H ₁₀ N ₂ O	6-HYDROXYHARMANE	

LIST OF SPECTRA INDICATE BY MW: 502-510				
TOTAL NO OF SPECTRA IN THE FILE: 1597				
ID #	MW	MF	NAMES	
1569	502	C ₁₀ H ₁₁ N ₂ O ₅	N(A)-CINNAMOYL-20-OH-CYLINDROCARINE	
575	504	C ₁₀ H ₁₁ N ₂ O ₄	0-(3,4-DIMETHOXYBENZOYL)VOACANGINOL	
1570	504	C ₁₀ H ₁₁ N ₂ O ₅	N(A)-DI-H-CINNAMOYL-20-OH-CYLINDROCARINE	
892	510	C ₁₂ H ₁₄ N ₄ O ₂	N(B')-CO2ET-TETRA-H-17R-USAMBARENSINE	

表 2 RLIST 目录

NAME SEARCH: CANTHIN-6-ONE				
TOTAL NO OF SPECTRA IN THE FILE: 1597				
ID #	MW	MF	NAMES	REF.
7	220	C ₁₁ H ₈ N ₂ O	CANTHIN-6-ONE	PLMEAA 46(3), 187(1982)
8	220	C ₁₁ H ₈ N ₂ O	CANTHIN-6-ONE	PLMEAA 36(2), 113(1979)
9	220	C ₁₁ H ₈ N ₂ O	CANTHIN-6-ONE	LLOYA2 46(3), 374(1983)

2) MWSER 和 MFSE——分子量检索和分子式检索:

当进入待检的 MW 或 MF 后,系统对全库谱数据作指纹检查,给出具有该分子量或分子式的全部图谱。当进入 MF 检索时系统提示要求按 C-H-N-O-S-Cl-Br-I 的顺序键入分子式,两部分结果分示于表 3、4。

表 3 MWSER 检索表

MW SEARCH: 642			
TOTAL NO OF SPECTRA IN THE FILE: 1597			
ID #	MW	MF	NAMES
627	642	C ₁₁ H ₁₆ N ₂ O ₂	MACROCARPAMINE
628	642	C ₁₁ H ₁₆ N ₂ O ₂	MACRALSTONIDINAL
911	624	C ₁₁ H ₁₆ N ₂ O ₂	MACROCARPAMINE
913	624	C ₁₁ H ₁₆ N ₂ O ₂	DECARBOMETHOXY--ANHYDROVOBTUSIN

表 4 MFSEER 检索表

MF SEARCH: C ₁₁ H ₁₆ N ₂ O ₂			
TOTAL NO OF SPECTRA IN THE FILE: 1597			
ID #	MW	MF	NAMES
627	642	C ₁₁ H ₁₆ N ₂ O ₂	MACROCARPAMINE
628	642	C ₁₁ H ₁₆ N ₂ O ₂	MACRALSTONIDINAL
911	642	C ₁₁ H ₁₆ N ₂ O ₂	MACROCARPAMINE
913	642	C ₁₁ H ₁₆ N ₂ O ₂	DECARBOMETHOXY--ANHYDROVOBTUSIN

3) MISER——特征离子检索:

本模块的主要工作内容和程序流程见表 5、图 2。

用户在屏幕提示下输入待检特征峰及设定以下两个参数:

a. “最低命中数”(MINIMUM NUMBER OF FITS), 系指“未知谱”与库存谱比较时, 被匹配上的峰个数必须等于或大于此数, 才能认为两谱“一致”。这一参数值的大小, 决定着检索的专一程度的高低。

b. “峰强因子”(INTENSITY FACTOR 1—10), 为了减少质谱峰强波动给检索带来的影响, 用一个“峰强因子”与未知谱中各峰的强度相乘或除, 即: 待检峰强/因子 < 被检峰强 < 待检峰强 × 因子。如: 因子为 2, 某待检峰的强度为 20%, 则库存谱中凡强度在 10~40% 的相应峰均被认为是“一致”的。峰强因子一般指定为 2。当限定被检峰强的上限(待检峰强 × 因子)大于 100% 时, 则令其等于 100% 处理。

表 5 MISER 检索表

MASS-INT SEARCH: 121, 144, 265			
TOTAL NO OF SPECTRA IN THE FILE: 1597			
ID #	MW	MF	NAMES
665	702	C ₁₁ H ₁₆ N ₂ O ₂	O-ACETYL-UMBELLAMINE
121(29),	144(11),	265(7).	
944	688	C ₁₁ H ₁₆ N ₂ O ₂	3R-AC-EPI-ERVAFOLIDENE
121(80),	144(70),	265(66).	
947	688	C ₁₁ H ₁₆ N ₂ O ₂	3S-AC-2,16-DIHYDRO-ERVAFOLIDENE
121(58),	144(78),	265(90).	

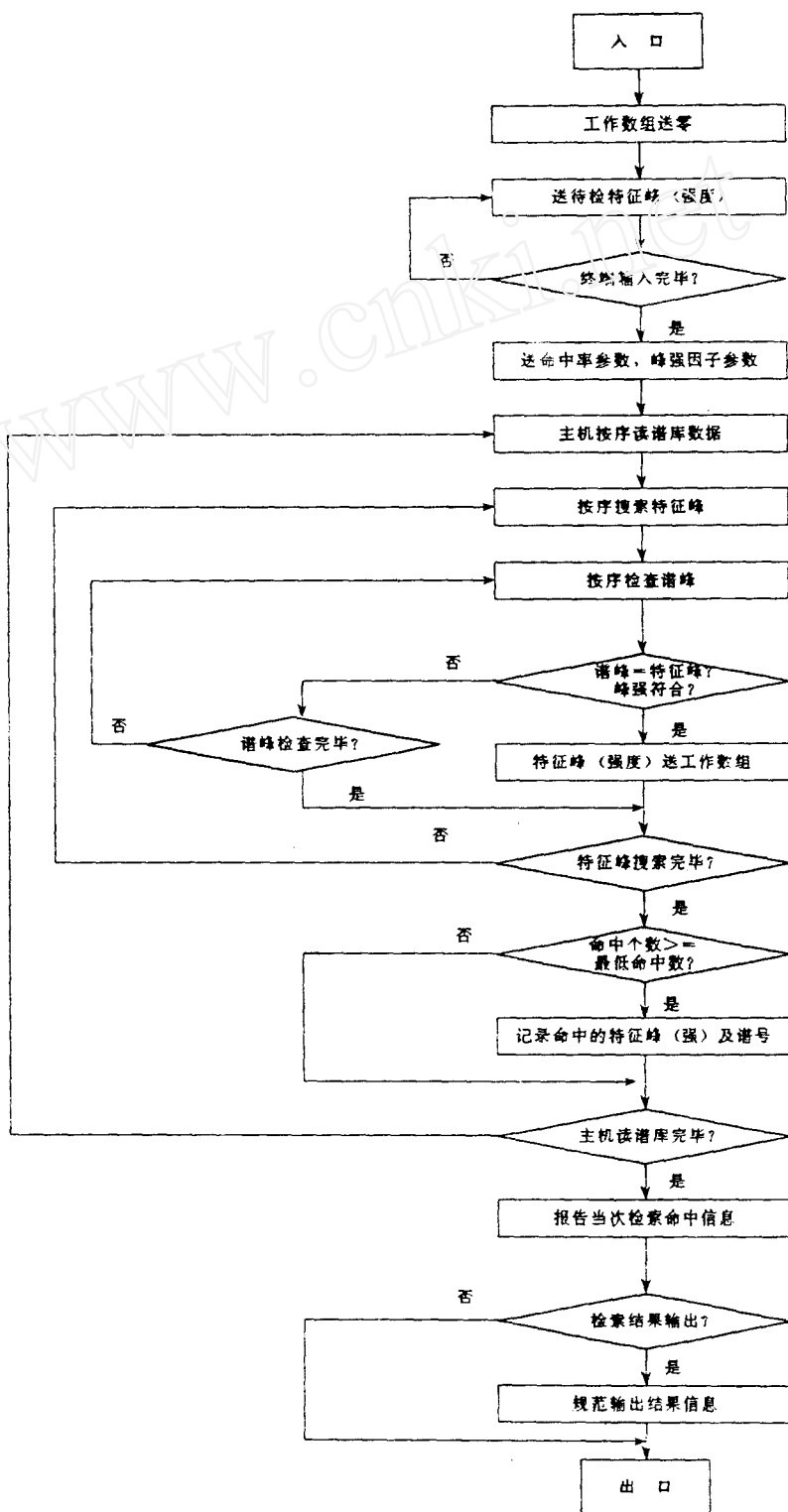


图2 特征离子峰检索模块程序流程图

4) NLSER——中性碎片检索

本模块用于检查从分子离子 M 到 m/z 150 以上范围内的一级中性丢失,系检查较为邻近 M 的高质量区中的中性碎片 NL,从中获取有关化合物的结构信息,当用户指定“最低命中数(1—10)”,并逐个输入待查的中性碎片(NL)数值后,模块程序则逐个搜索谱库数据所产生的(NL)数值,并作命中记录。部分检索结果及程序流程图见表 6、图 3。

表 6 NLSER 检索表

NEUTRAL LOSS SEARCH: 17, 29, 73			
TOTAL NO OF SPECTRA IN THE FILE: 1597			
ID#	MW	MF	NAMES
63	294	$C_{10}H_{22}N_2O$	TETRADEHYDRO—18,19—2H—CORYNANTHOL
17(13),	29(14),	73(57).	
94	298	$C_{10}H_{22}N_2O$	18,19—DIHYDROANTIRHINE
17(2),	29(4),	73(45).	
139	310	$C_{20}H_{24}N_2O$	21-DESOXY-AJMALINE
17(7),	29(5),	73(7).	
284	352	$C_{21}H_{24}N_2O_2$	19—HYDROXY—CORONARIDINE
17(15),	29(23),	73(13).	

* 系统列出命中的中性碎片值与 MW—NL 峰的(强度),以便使用户了解发生这一中性碎征丢失的机率大小。

5) MLIST—图谱的质量表的显示与输出:本模块完成图谱的质量表显示或输出见表 7。

表 7 MLIST—图谱质量表

SPEC. NO:624 MW:634 MF: $C_{17}H_{14}N_4O_2$ NAME: BISNORDIHYDROTOXIFERINE			
PEAK	MASS	I/BASE(%)	
1	635	29	
2	634	95	
3	633	35	
4	437	37	
5	345	47	
6	210	22	
7	144	53	
8	108	100	

6) BPLOT——棒状图输出:本功能块具有完整的谱图绘制功能,如:

选择棒图及质量表的输出方式。

选择棒图的输出形状。

峰强的放缩功能。

图形的比例放缩及彩色调配功能。

这些功能可以满足谱集出版及幻灯片制作等需要。部分输出结果及模块程序流程见图 4、5。

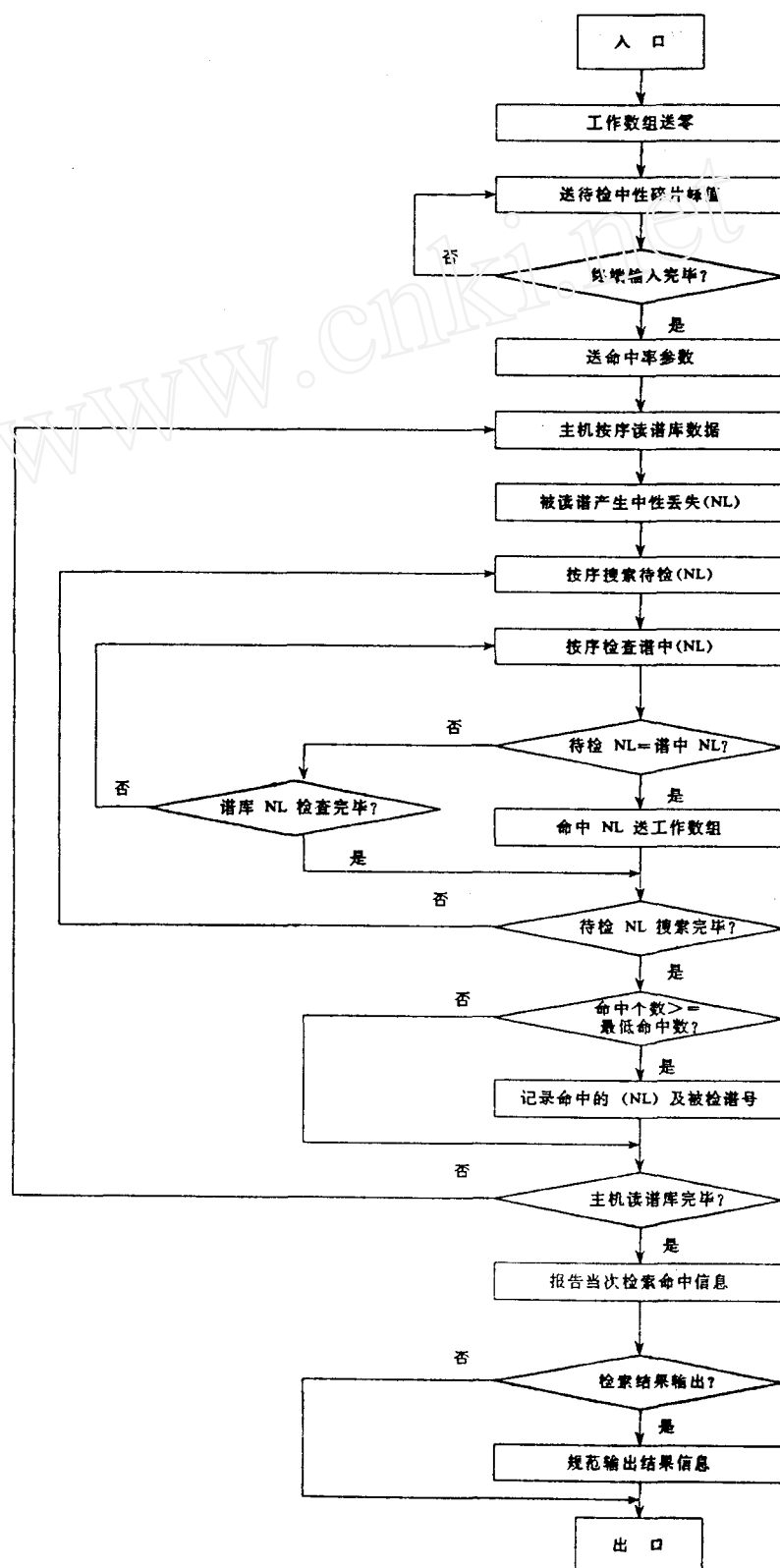
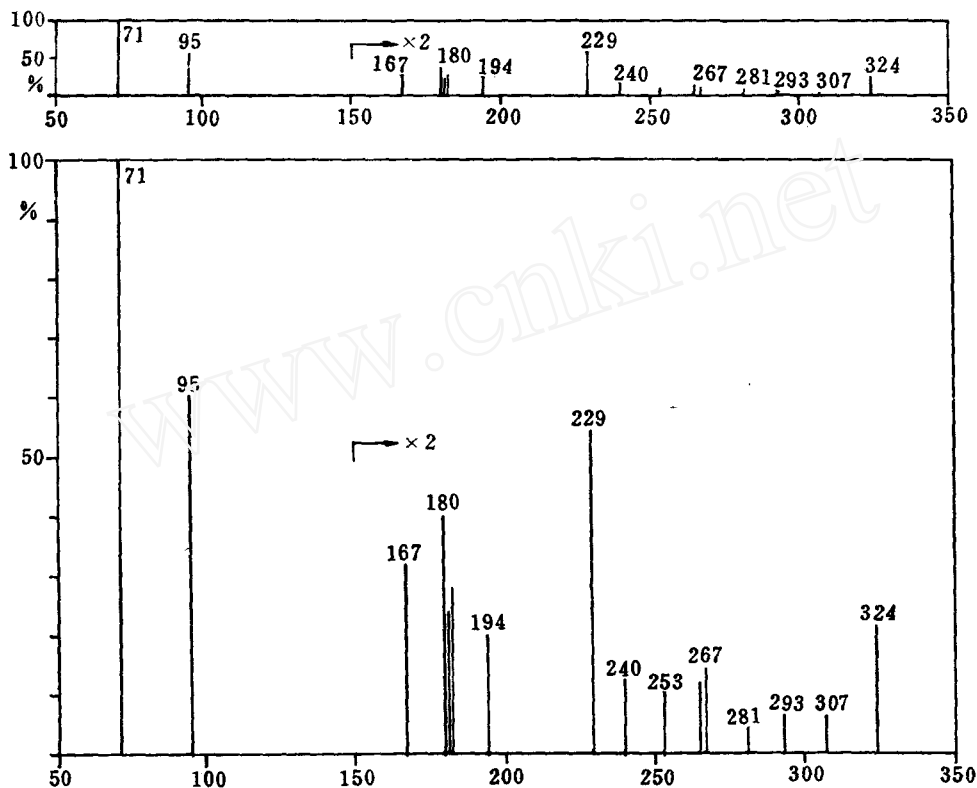


图3 中性碎片检索模块程序流程图



SIMM-SYSTEM

MASS	INT.	MASS	INT.
71	100	240	6
95	60	253	5
167	16	265	6
180	20	267	7
181	12	281	2
182	14	293	3
194	10	307	3
229	27	324	11

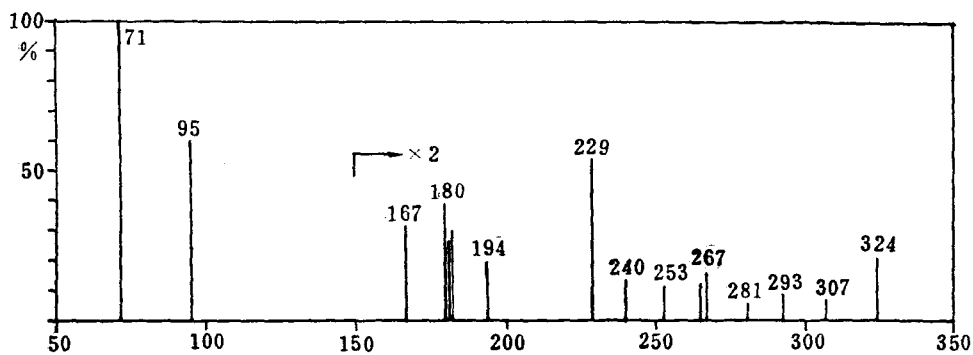


图4 棒图输出功能

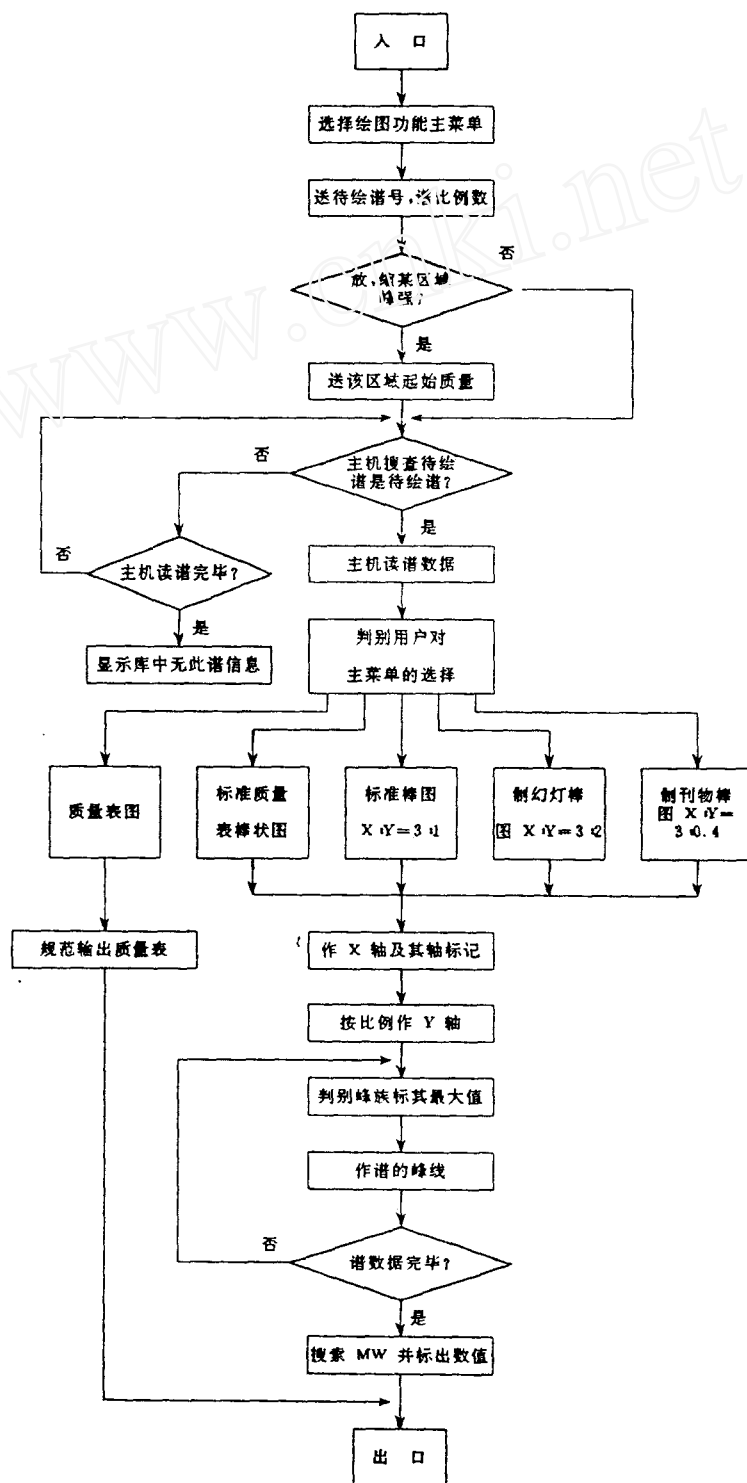


图 5 绘图功能模块程序流程图

四、应用举例

1. N-CARBOMETHOXY-11-HYDROXY-12-METHOXYKOPSINALINE (1); 本品为自云南蕊木(KOPSIA OFFICINALIS)的果实中分得的一个吲哚生物碱, 分子量 458 ($C_{24}H_{30}N_2O_7$), 其质谱如图 6 所示:

Unknown (1)

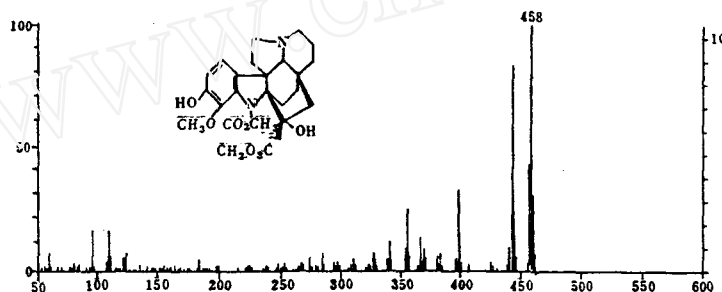


图 6 “未知物”(1)的质谱

首先做分子量检索, 得到答案 5 个(表 8)。

表 8 未知物(1)分子量检索

MW SEARCH: 458			
TOTAL NO OF SPECTRA IN THE FILE: 1597			
ID #	MW	MF	NAMES
567	458	$C_{24}H_{30}N_2O_7$	DIHYDROKOPSINGINE
1085	458	$C_{24}H_{30}N_2O_7$	N-CO ₂ ME-11-MEO-12-OH-KOPSINALINE
1091	458	$C_{24}H_{30}N_2O_7$	N-CO ₂ ME-11-OH-12-MEO-KOPSINALINE
1558	458	$C_{25}H_{32}N_2O_6$	DIHYDROVINDOLINE
1559	458	$C_{25}H_{34}N_2O_6$	14,15-DIHYDROVINDOLINE

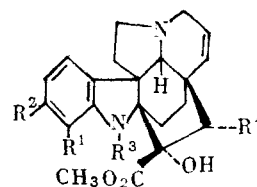
继续用特征峰检索, 选取特征质量 9 个—m/z 458, 457, 443, 339, 367, 356, 275, 124, 109, 它们代表了高质量区与低质量区的主要碎片。将 m/z 值连同丰度同时键入, 规定峰强因子为 2, 命中 3 个以上峰值的检索结果见表 9。

表 9 特征峰检索结果

MASS—INT SEARCH,						
TOTAL NO OF SPECTRA IN THE FILE, 1597						
ID#	MW	MF	NAMES			
470	384	C ₂₂ H ₂₂ N ₂ O ₄	(-)-12-METHOXYKOPSINALINE			
109(99)	124(99)	356(99)				
567	458	C ₂₄ H ₃₀ N ₂ O ₇	DIHYDROKOPSINGINE			
109(99)	124(99)	399(99)	443(99)	453(99)		
569	472	C ₂₅ H ₃₂ N ₂ O ₇	(-)-N-CO2ME-11,12-DI-OME-KOPSINALINENE			
109(99)	124(99)	457(99)				
1077	472	C ₂₅ H ₃₂ N ₂ O ₇	N-CARBOMETHOXY-11,12-DIMETHOXY-KOPSINALENE			
109(20)	124(11)	457(54)				
1085	458	C ₂₄ H ₃₀ N ₂ O ₇	N-CO2ME-11-MEO-12-OH-KOPSINALINE			
109(16)	124(8)	356(34)	367(17)	399(28)	457(56)	458(100)
1091	458	C ₂₄ H ₃₀ N ₂ O ₇	N-CO2ME-11-OH-12-MEO-KOPSINALINE			
109(17)	124(7)	356(27)	399(34)	443(84)	457(46)	458(100)

表 10 基本结构特征

R ¹	R ²	R ³	R ⁴	COMPOUND
CH ₃ O	H	H	H	# 470
CH ₃ O	H	CO ₂ CH ₃	OH	# 567
CH ₃ O	CH ₃ O	CO ₂ CH ₃	H	# 569 = # 10
HO	CH ₃ O	CO ₂ CH ₃	H	# 1085
CH ₃ O	HO	CO ₂ CH ₃	H	# 1091



1085 与 # 1091 与“未知物(1)”的质谱具有最相似的谱形,所有的“答案”均具有 A—B—C—D—E—F 环系统的结构。这是仅依靠低分辨质谱数据而未加其它光谱的信号与理化常数,检索结果给出基本结构特征,达到极为近似的程度。值得指出,本系统收录了少数不完整的质谱数据,在文献记载中,只有特征峰的质量而不附峰强值(本系统对这些峰的强度指定为 99%)。由本例可以看出,这些不完全的谱也可以提供有相当参考价值的结构信息(表 10)。

2. TABENSONINE N—OXIDE (2) 由思茅山橙(MELODINUS HENRYI CRAIL)中分得。未知物(2), MW 352 ($C_{21}H_{24}N_2O_3$),其质谱示于图 7。

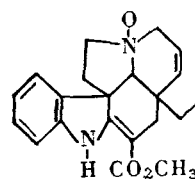
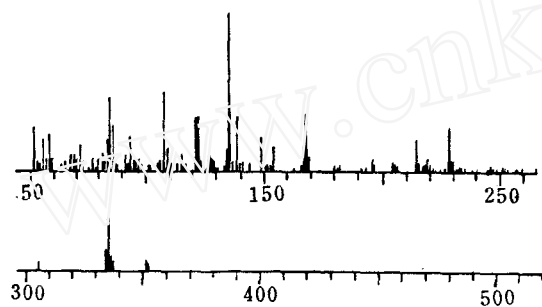
分子量检索,得答案 80 个,经逐一核对,证明谱库中没有相同的谱,很可能为新化合物。为了获得与“未知物”近似化合物的结构信息作为解析图谱时的判断依据,我们选择了

除 MW 以外的特征峰 8 个 (m/z 336, 229, 228, 168, 135, 122, 121, 107) 输入 m/z -INT, 指定峰强因子为 2, 作特征峰检索, 令系统输出“命中”5 个峰以上的答案, 见表 11。

表 11 特征峰检索结果

MASS-INT SEARCH,					
TOTAL NO OF SPECTRA IN THE FILE, 1597					
ID#	MW	MF	NAME		
57	292	$C_{10}H_{14}N_4$	5,18-SECO-DIENE		
107(99)	121(99)	122(99)	135(99)	168(99)	
231	336	$C_{21}H_{24}N_2O_2$	TABERSONINE		
121(31)	122(40)	135(100)	168(24)	229(24)	336(55)
233	336	$C_{21}H_{24}N_2O_2$	16-EPI-19S-VINDOLININE		
121(24)	122(23)	135(94)	229(28)	336(57)	
934	808	$C_{28}H_{38}N_2O_4$	ISOVINCATHICINE		
107(99)	121(99)	122(99)	135(100)	168(99)	
943	672	$C_{22}H_{28}N_2O_4$	2,16-DIHYDRO-N-AC-ERVAFOLINE		
107(99)	212(99)	122(99)	135(99)	168(99)	
944	688	$C_{22}H_{28}N_2O_5$	3R-AC-EPI-ERVAFOLIDENE		
107(99)	121(99)	122(99)	135(99)	168(99)	228(99)
945	690	$C_{22}H_{28}N_2O_5$	2,16-DIHYDRO-3R-AC-EPI-ERVAFOLIDENE		
107(99)	121(99)	122(99)	135(99)	168(99)	
947	688	$C_{22}H_{28}N_2O_5$	3S-AC-2,16-DIHYDRO-ERVAFOLIDENE		
107(99)	121(99)	122(99)	135(99)	168(99)	228(99)
1483	336	$C_{21}H_{24}N_2O_2$	(+) -PSEUDOTABERSONINE		
107(99)	121(99)	122(99)	135(99)	168(99)	228(99) 336(99)
1485	336	$C_{21}H_{24}N_2O_2$	VENALSTONINE		
121(99)	122(99)	135(99)	229(99)	336(99)	
1587	672	$C_{22}H_{28}N_2O_4$	3-O-ACETYL-EPI-ERVAFOLIDENE		
107(99)	121(99)	122(99)	135(99)	168(99)	228(99)
1588	672	$C_{22}H_{28}N_2O_4$	3-O-ACETYL-ERVAFOLIDENE		
107(99)	121(99)	122(99)	135(99)	168(99)	228(99)
1589	690	$C_{22}H_{28}N_2O_5$	3-O-ACETYL-ERVAFOLIDENE		
107(99)	121(99)	122(99)	135(99)	168(99)	

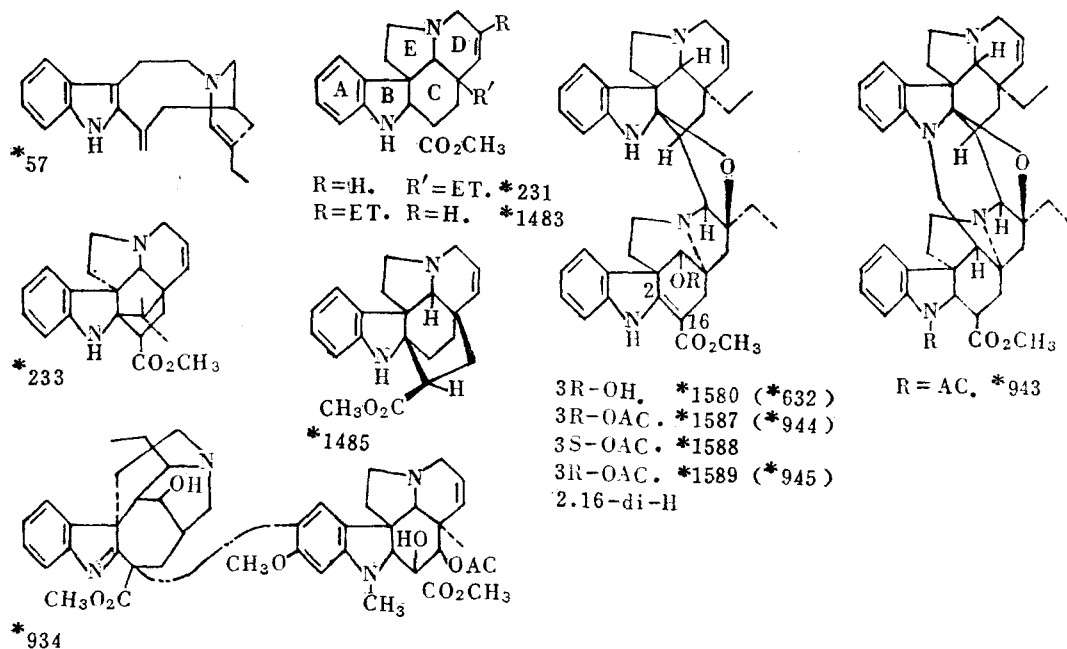
13 组答案中,有 3 组为重复的库存谱,共有 11 种结构,内除 1 种外,其余 10 种(包括 6 种双吲哚生物碱)全属 Vincadifformine 类型,仅在不饱和键与取代基的位置有所差异。最终,结合其它光谱与理化数据,鉴定该未知物为 TABERSONINE N-OXIDE。结合在参考谱不存在于谱库中时,类似的有关数据对未知物结构推测,往往能提出有用的指示(表 12)。



Unknown (2)

图 7 “未知物”(2)的质谱

表 12 结构推测



参 考 文 献

- [1] M. Hesse, *Alkaloid Chemistry*, P11, Georg Thieme Verlag (1978)
[2] M. Hesse, *Indolalkaloide*, Teil 1 u. 2, Verlag Chemie (1974)
[3] B. Gabetta, G. Mustich, *Spectral Data of Indole Alkaloids*, Inverni Della Zeffa, Milan (1975)

A Computer System for Searching the Mass Spectra of Indole Alkaloids

L. Z. Hsueh, L. Y. Huang, W. Z. Xu*, Y. L. Zhou** and Z. H. Huang
(Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai)

Received 30, Aug. 1988

Abstract

A small-scale, home-made computer system for searching the mass spectra of indole alkaloids has been built up which consists of a data file and a set of searching subroutines. Included in this system are;

(1) A data file involving 1600 mass spectra of indole alkaloids taken from literature and (partly) measured in our laboratory in the abbreviated form after Biemann, i. e. , picking 2 peaks out of every 14-mass interval within the whole range of the spectrum.

(2) A search program capable of executing fingerprint match as well as spectra comparison approached in different manners; molecular weight, molecular formula, selected ion peaks and selected neutral losses.

(3) The out-put of results is given in the form of numerical tabulation or bargraphs as requested by the user. The data file can be modified easily by deleting or renewing modules.

* Computing Center, Shanghai Branch, Chinese Academy of Sciences

** Author to whom correspondence should be addressed