

# 采用KnowItAll® 软件进行多种光谱技术联用检索

Gregory Banik, Ph.D., Ty Abshear, and Karl Nedwed

Bio-Rad Laboratories, Inc., Informatics Division, Philadelphia, PA, USA

## Spectroscopy

280060

### 摘要

核磁共振 (NMR)，质谱 (MS) 以及红外光谱 (IR) 是长期以来进行未知物解析、化合物确证以及结构剖析的工具。通常，匹配系数 (HQI) 用于描述一个未知物的谱图和已知物标准谱图的相关性。本文推出一种新的软件系统，通过使用多种互补的光谱技术对化合物进行解析，从而避免使用单一技术可能产生的不确定性。

### 背景

各种光谱和色谱技术 (图1) 应用于解决这些基本的问题：化合物确证和未知物鉴定。前者主要用于质控管理，而后者主要用于化合物合成及天然产物鉴定。

技术	获取的相关信息
红外/拉曼	所代表的官能团信息
核磁共振	碳原子和氢原子的键合
质谱	分子的质量以及所代表的原子
紫外-可见	分子的电荷系统状态
色谱	样品的组成成分及其相关性

图1：通过相关分析技术获取的信息

随着计算机检索技术的发展，化合物确证和未知物鉴定方法得到了明显的改善。计算出的未知物谱图和数据库中参考谱图的对比可以为其鉴定提供强有力的证据，或者说至少为最终的确定或确证提供了线索。

在检索时，要想处理好结果与参考谱图的对比及相关的化学结构，我们通常要考虑三个基本的概念：

1. 匹配表 - 软件程序在检索后生成一系列相似的谱图组成的表格。

2. 匹配系数 (HQI) - 用于衡量各个参考谱图与未知物谱图相关性的数学描述。未知谱图与参考谱图相似度越高，则匹配系数越高。
3. 排序 - 匹配表将会根据匹配系数进行排序以确保具有高匹配系数的谱图位于前列，而匹配系数低的将列于末尾。

这三个概念可以通过图2说明，该图描述了一个红外光谱检索的排序表。

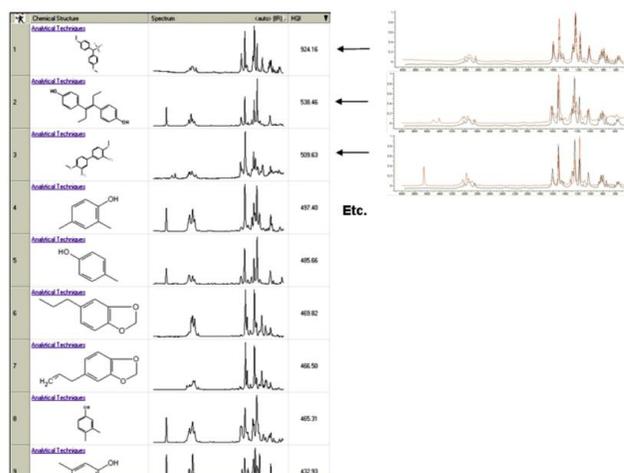


图2：谱图匹配列表及匹配系数 (HQI)

### 多种技术联用检索 – 原理及实践

Bio-Rad Laboratories, Inc.推出了强大的多技术联用检索的系统。检索的过程非常简洁：通过检索不同光谱技术的数据库、集合各种互补的光谱技术的结果来解析未知物。它有两项创新把传统的匹配表转换为易于识别的结果。

第一项创新是将单一光谱的检索结果用数字形式排列。如图3所示，这项技术将命中谱图按相对匹配排列。第二项创新则更有意义，它转换不同技术的匹配表为空间散点。每一个位于N维光谱匹配系数坐标空间的点代表了一个单一的化合物（图4）。越靠近原点的化合物和未知物的谱图相似性越低。

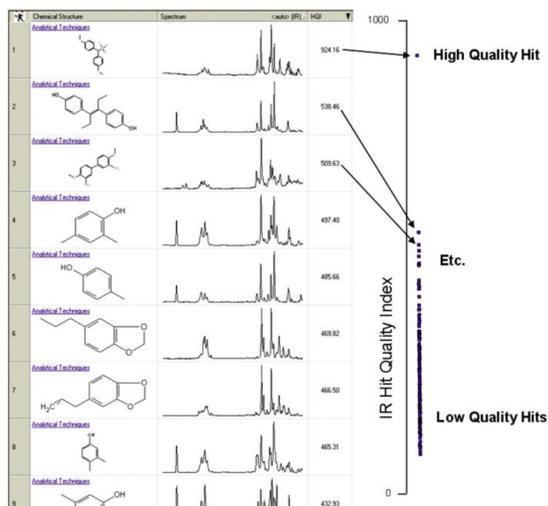


图3 – 通过单技术检索得到的一维匹配系数散点

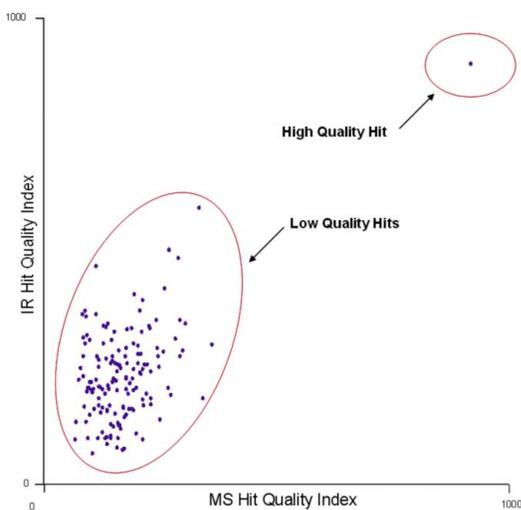


图4 – 通过两种技术检索得到的匹配系数散点

### 资源与方法

- 软件: KnowItAll® Informatics System
  - 红外检索 – 欧氏距离算法
  - 质谱检索 – 修正的Finnigan算法

### 核磁共振谱 – 峰检索算法

- 红外数据库: KnowItAll® IR Spectral Library 红外谱库
- 核磁共振谱数据库: KnowItAll® NMR Spectral Library 核磁共振谱库
- 质谱数据库: KnowItAll® MS Spectral Library 质谱谱库
- 测试谱图集

- 红外、质谱 – NIST WebBook (<http://webbook.nist.gov>)
- 核磁共振谱图 – Bio-Rad样品文件

通过KnowItAll进行多种技术谱图检索（图5）。各种光谱技术的检索结果可以同时得到。不同技术结果通过化学结构关联，从而得到如图6所示的匹配系数散点。

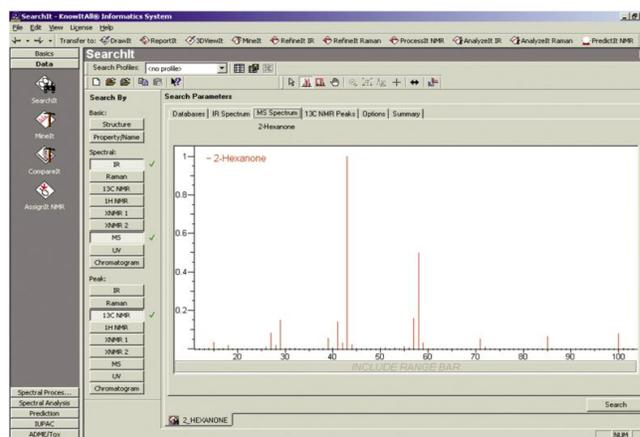


图5 – 定义多种技术联用检索

谱图必须要求在各种技术谱库中都存在，例如，通过红外/质谱联合检索得到的命中结果必须对应于同一个结构式。这一措施避免了红外光谱得到检索结果但质谱却没有，反之亦然。

为了避免NIST/EPA/NIH质谱库带来的对结果的影响，测试中并没有选择这些NIST WebBook中的谱图。

命中最前列200个参考谱库化合物匹配系数散点如图6。交互式显示允许用户通过点击单个数据点或套索一组数据点来对应所选化合物的其他信息。

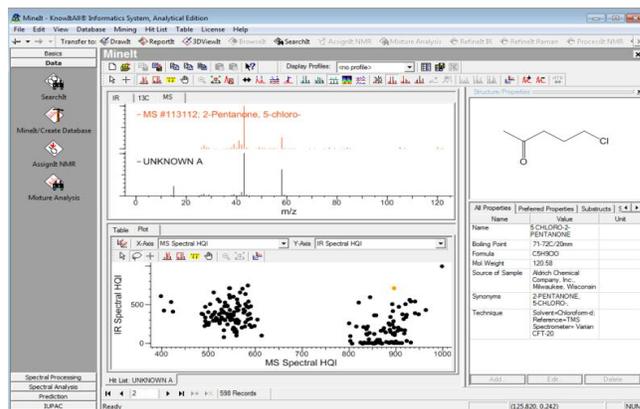


图6 – 多种技术检索匹配系数散点图（核磁/质谱/红外）

## 结果和讨论

### 例1 – 胆甾醇乙酸酯

这个例子通过对胆甾醇乙酸酯的红外和质谱图同时检索（图7），在谱库中得到精确的匹配结果。质谱和红外的二维匹配系数散点图显示，单独的红外或是质谱都不能确证此化合物（图8）。然而，通过集合两种技术，KnowItAll可以确定这个未知物就是胆甾醇乙酸酯。有趣的是，许多相关的胆固醇酯（包括胆固醇）类化合物都位于图表的右上方。这表明谱图空间和化学结构空间的高度相关性。

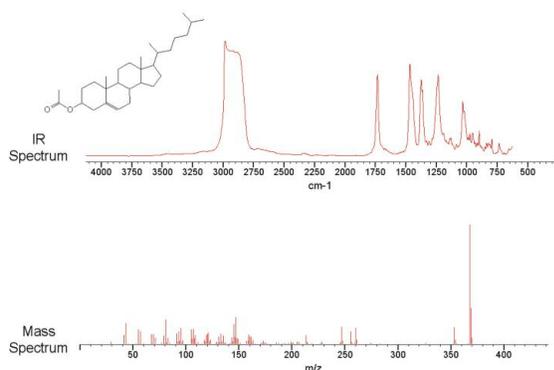


图7 – 例1: 胆甾醇乙酸酯

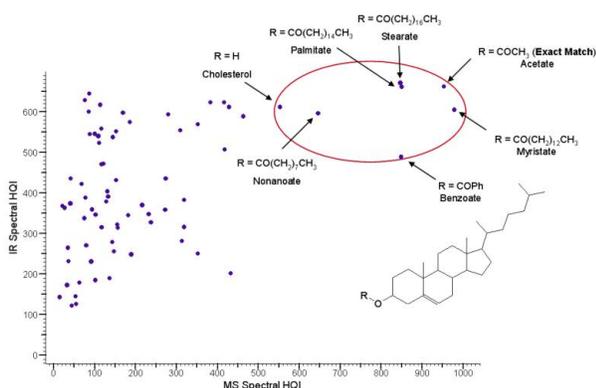


图8 – 例1 结果: 72 双谱图 (质谱/红外)

### 例2 – 1,1,4,4-四苯基-1,3-丁二烯

这个例子同时检索1,1,4,4-四苯基-1,3-丁二烯的红外和质谱图（图9），在相关的数据库中得到精确的匹配结果。由二维质谱与红外匹配系数散点图可以看出，单一的红外或者质谱并不能对此化合物进行确证（图10）。然而，通过并行的两种技术，KnowItAll可以明确这个谱图就是1,1,4,4-四苯基-1,3-丁二烯。

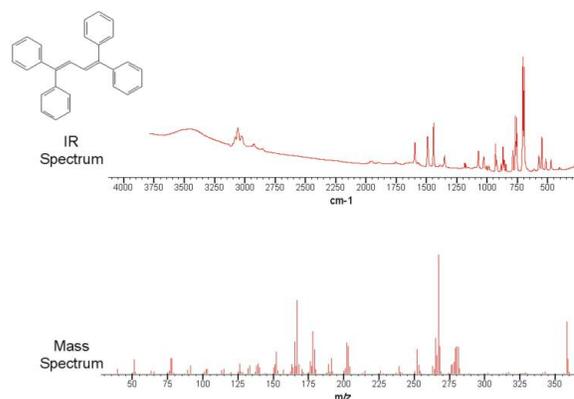


图9 – 例2: 1,1,4,4-四苯基-1,3-丁二烯

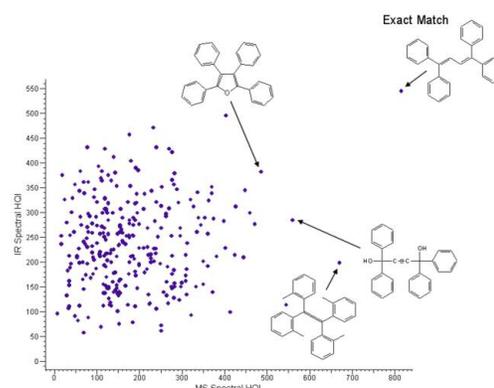


图10 – 例2 结果: 287 双谱图 (质谱/红外)

### 例3 – 邻苯二甲酸丁基环己基酯

在这个例子中，邻苯二甲酸丁基环己基酯（图11）的红外和质谱图同时检索没有在数据库中得到精确的匹配。然而，质谱和红外的二维匹配率图表（图12）却展现了一个有趣的现象：在图表的右上方堆积了一组化合物。他们是11种邻苯二甲酸二酯化合物。由此我们再一次看到谱图空间和化合物结构空间的紧密关联。

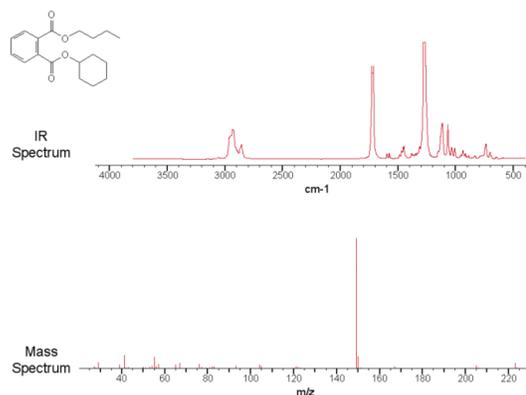


图11 – 例3: 邻苯二甲酸丁基环己基酯

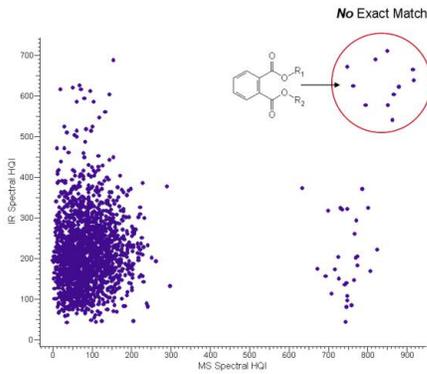


图12 - 例3结果: 1821双谱图 (质谱/红外)

#### 例4 - 甲丁酮

这个例子同时检索甲丁酮 (图13) 的碳谱、红外和质谱图, 从相关的数据库中得到精确的匹配。二维碳谱与红外匹配系数散点 (图14), 二维质谱与碳谱的匹配系数散点 (图15) 及二维质谱与红外的匹配系数散点 (图16) 表示, 单一的核磁、红外或者质谱并不能对此化合物进行确证。然而, 联合三种技术, KnowItAll可以无疑地确定这个物质就是甲丁酮。再次强调, 散点图右上方的聚集相似化合物, 表明谱图空间与化学结构空间的密切相关。

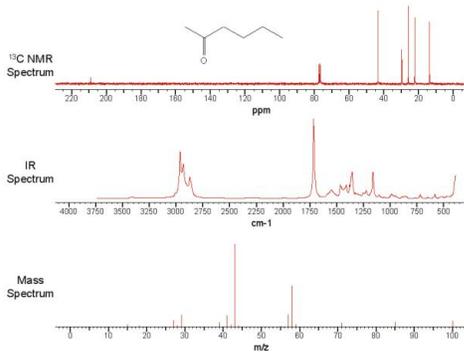


图13 - 例4: 甲丁酮

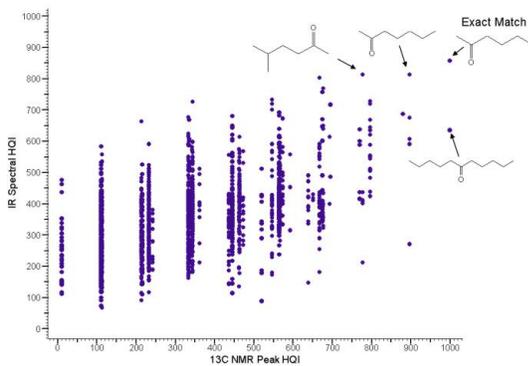


图14 - 例4结果: 5715组谱图 (核磁/红外)

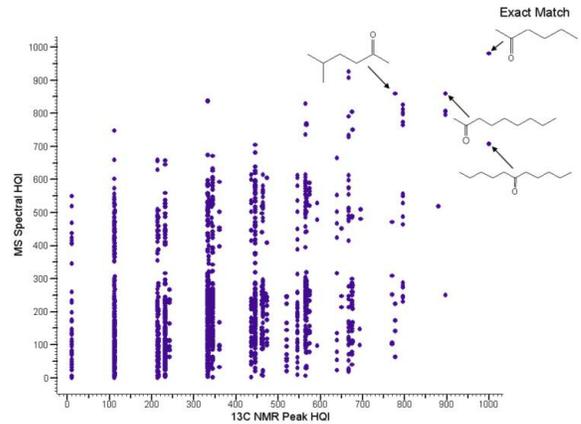


图15 - 例4结果: 5715组谱图 (核磁/质谱)

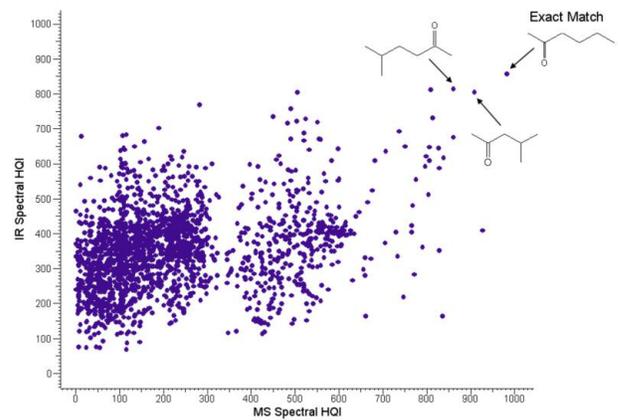


图16 - 例4结果: 5715组谱图 (质谱/红外)

#### 结论

以上研究给我们这个快速易用的多技术联用检索系统如下结论:

- 多种技术谱图检索分析方法简单明了, 光谱空间变得直观。
- 这项发明有传统列表不具备的直观, 使它成为有效的数据挖掘工具。
- 显然, 这项技术适用于化合物确证、未知物鉴定以及结构剖析。



Bio-Rad  
Laboratories, Inc.