Structural Analysis of Small Molecules Using Precursor Ion Fingerprinting (PIF)

Robert Mistrik
HighChem Ltd.
Bratislava, Slovakia

www.highchem.com

CoSMoS Boston 2009
Overview

Traditional Techniques

• Spectra search against reference libraries
• Spectra classification
• Accurate mass search
• Predictive fragmentation

New Method

• Precursor Ion Fingerprinting (PIF)
Mass Spectrometry is an indirect structure elucidation method.

From mass to shape:

From mass to arrangement:

<table>
<thead>
<tr>
<th>m/z</th>
<th>147.03</th>
<th>183.03</th>
<th>201.08</th>
<th>202.14</th>
<th>211.05</th>
<th>227.09</th>
<th>228.37</th>
<th>229.10</th>
</tr>
</thead>
</table>

Chemical structure with m/z values.
“A mass spectrum of an organic molecule exhibits a peak for almost each mass which can conceivably be constructed from any number of the atoms present in the molecule”

Fragmentation Processes

[Diagram showing various molecular structures and their mass-to-charge ratios (m/z) with arrows indicating fragmentation processes.]
The primary goal of spectra classification is to find correlation between the properties of compounds and their mass spectra: common substructures, point of origin, toxicity, smell, color, taste etc.

![Classification Diagram]

### Classification Method

<table>
<thead>
<tr>
<th>Class A</th>
<th>Class B</th>
<th>Class X</th>
</tr>
</thead>
</table>

### Binary Classifiers

- NO
- NO
- YES
Each spectrum is represented as a single point in an n-dimensional space.
Accurate mass search against structural libraries

Mass distribution of 7408 endogenous metabolites
(curated KEGG Ligand Database structures)

Unit resolution

\[ \Delta(m/z) = 0.00001 \]
Immense Number of Structural Isomers

MS$^3$ Spectrum: ESI (+), LTQ Orbitrap
Formula Generator: Mass Frontier 6.0
Structure Isomers: Molgen 3.0
Consistency checking between library structure and mass spectrum

Database of known human metabolites

Automated of Prediction Fragments

Unknown Component in Biofluid
Metabolite ID using Predictive Fragmentation

MS² of Irinotecan (parent drug)

- m/z 587.28641
- m/z 543.29658
- m/z 587.28693

MS² of Irinotecan metabolite (M6)

- m/z 603.28133
- m/z 559.29150
- m/z 597.28641
- m/z 559.29134

HighChem
Precursor Ion Fingerprinting (PIF)  
A new concept for the interpretation of mass spectra

Efficient structure characterization of small molecules remains a huge bottleneck in an array of disciplines.

Principle of the structure elucidation using PIF

1. Mass Spectrum
2. Substructures
3. Structure Assembly
4. Structure Candidate(s)
5. Precursor Ion Fingerprinting (PIF)
Quasi-equilibrium theory: The probabilities of the various possible decomposition products of an ion depend upon its structure, internal energy, and the energy deposited during ion activation, but are independent of the structure of the precursor ion and the formation mechanism of the ion undergoing decomposition.
Quasi-equilibrium theory:

“The probabilities of the various possible decomposition products of an ion depend upon its structure, internal energy, and the energy deposited during ion activation, but are independent of the ionization method, the structure of the precursor ion, and the formation mechanism of the ion undergoing decomposition.”

Precursor Ion Fingerprinting (PIF)

ESI (+), Finnigan LCQ
Precursor Ion Fingerprinting (PIF)

ESI (-), Finnigan LCQ
Prediction of Fragmentation Pathways
# Mass Frontier Fragmentation Library™

<table>
<thead>
<tr>
<th>Total number of</th>
<th>Mass Frontier 5.0</th>
<th>Mass Frontier 6.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fragmentation Schemes</td>
<td>20.003</td>
<td>30.936</td>
</tr>
<tr>
<td>Individual Reactions</td>
<td>87.966</td>
<td>129.229</td>
</tr>
<tr>
<td>Chemical Structures</td>
<td>101.850</td>
<td>151.762</td>
</tr>
<tr>
<td>Decoded Mechanisms</td>
<td>81.389</td>
<td>120.029</td>
</tr>
<tr>
<td>Source</td>
<td>Volume</td>
<td>Year</td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>--------</td>
<td>----------</td>
</tr>
<tr>
<td>JASMS (Journal of the American Society for Mass Spectrometry)</td>
<td>1-17</td>
<td>1990-2006</td>
</tr>
<tr>
<td>RCM (Rapid Communications in Mass Spectrometry)</td>
<td>1-20</td>
<td>1987-2006</td>
</tr>
<tr>
<td>JMS (Journal of Mass Spectrometry)</td>
<td>30-41</td>
<td>1995-2006</td>
</tr>
<tr>
<td>OMS (Organic Mass Spectrometry)</td>
<td>1-29</td>
<td>1968-1994</td>
</tr>
<tr>
<td>MSR (Mass Spectrometry Reviews)</td>
<td>1-25</td>
<td>1981-2006</td>
</tr>
<tr>
<td>EJMSBMER (European Journal of Mass Spectrometry in Biochemical, Medicine, and Environmental Research)</td>
<td>1-2</td>
<td>1980-1982</td>
</tr>
<tr>
<td>BMS (Biomedical Mass Spectrometry)</td>
<td>1-12</td>
<td>1974-1985</td>
</tr>
<tr>
<td>BEMS (Biomedical and Environmental Mass Spectrometry)</td>
<td>14-19</td>
<td>1987-1990</td>
</tr>
<tr>
<td>BMS (Biological Mass Spectrometry)</td>
<td>20-23</td>
<td>1991-1994</td>
</tr>
<tr>
<td>EJMS (European Journal of Mass Spectrometry)</td>
<td>4</td>
<td>1998</td>
</tr>
</tbody>
</table>
User provided fragmentation mechanism can improve fragment prediction for metabolites

User suggested fragmentation mechanism for parent drug

User’s mechanism applied to predicted metabolite structure
Automated Interpretation of Mass Spectra

- Library Search
- Matching product-ion spectra in PIF Library
- Identified fragments/substructures
- Structural constraints
- Consistency Checking
- Structural proposal
- Structure generator

Product-ion spectra of unknown

Library Search

Consistency Checking

structure generator

Patent pending
ESI (+), Finnigan LCQ
**Bad news:** Success rate of ion identification depends on the comprehensiveness of the reference library of ion spectra (PIF Library).

**Good news:** Metabolite ID operates in a structurally finite space!
Metabolite ID using PIF
Talinolol

EMS m/z 380

Spectral Tree of Talinolol
Fragment structure assignment

- Fragmentation reactions of parent drug stored in fragmentation library
- Applied to prediction of fragments for parent drug metabolites

Fragmentation Library™:
- HighChem library 100,000 mechanisms
- User defined mechanism

Parent drug (Library)

Metabolite (Prediction)
Matching product-ion spectra (MS³) reveal common substructures of parent drug and metabolites.
<table>
<thead>
<tr>
<th>Metabolite Structure Assignment</th>
<th>m/z 209</th>
<th>m/z 226</th>
<th>m/z 308</th>
<th>m/z 324</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Talinolol</strong></td>
<td><img src="image1" alt="Talinolol" /></td>
<td><img src="image2" alt="Talinolol" /></td>
<td><img src="image3" alt="Talinolol" /></td>
<td><img src="image4" alt="Talinolol" /></td>
</tr>
<tr>
<td><strong>Metabolite C01</strong></td>
<td><img src="image5" alt="Metabolite C01" /></td>
<td><img src="image6" alt="Metabolite C01" /></td>
<td><img src="image7" alt="Metabolite C01" /></td>
<td><img src="image8" alt="Metabolite C01" /></td>
</tr>
<tr>
<td><strong>Metabolite C02</strong></td>
<td><img src="image9" alt="Metabolite C02" /></td>
<td><img src="image10" alt="Metabolite C02" /></td>
<td><img src="image11" alt="Metabolite C02" /></td>
<td><img src="image12" alt="Metabolite C02" /></td>
</tr>
<tr>
<td><strong>Metabolite C09</strong></td>
<td><img src="image13" alt="Metabolite C09" /></td>
<td><img src="image14" alt="Metabolite C09" /></td>
<td><img src="image15" alt="Metabolite C09" /></td>
<td><img src="image16" alt="Metabolite C09" /></td>
</tr>
</tbody>
</table>

Intensity differences due to different hydroxyl group position.
PIF concept allows high throughput, semi-automated qualitative analysis of structurally related compounds (xenobiotic or endogenous metabolites)

Data provided by G. Hopfgartner, University of Geneva, Switzerland
Metabolite ID using PIF
Naltrexone

TIC full scan

Metabolite ID using PIF
Naltrexone

TIC full scan

Metabolite ID using PIF
Naltrexone

TIC full scan

Metabolite ID using PIF
Naltrexone

TIC full scan
Prediction of fragmentation pathways of Naltrexone using HighChem Mass Frontier and Fragmentation Library™

Naltrexone

m/z 342

m/z 324

m/z 282

HighChem Fragmentation Library™

ID: 6.735 of 89.326

Metabolite ID using PIF
Naltrexone

Site of glucoronization can be determined using fragmentation prediction and PIF
<table>
<thead>
<tr>
<th>Metabolite Structure</th>
<th>m/z 267</th>
<th>m/z 270</th>
<th>m/z 282</th>
<th>m/z 326</th>
<th>m/z 344</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Metabolite</strong></td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
</tr>
<tr>
<td><strong>Metabolite I</strong></td>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
</tr>
<tr>
<td><strong>Metabolite II</strong></td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
<td><img src="image13.png" alt="Image" /></td>
<td><img src="image14.png" alt="Image" /></td>
<td><img src="image15.png" alt="Image" /></td>
</tr>
<tr>
<td><strong>Metabolite III</strong></td>
<td><img src="image16.png" alt="Image" /></td>
<td><img src="image17.png" alt="Image" /></td>
<td><img src="image18.png" alt="Image" /></td>
<td><img src="image19.png" alt="Image" /></td>
<td><img src="image20.png" alt="Image" /></td>
</tr>
<tr>
<td><strong>Metabolite IV</strong></td>
<td><img src="image21.png" alt="Image" /></td>
<td><img src="image22.png" alt="Image" /></td>
<td><img src="image23.png" alt="Image" /></td>
<td><img src="image24.png" alt="Image" /></td>
<td><img src="image25.png" alt="Image" /></td>
</tr>
</tbody>
</table>
Metabolite ID using PIF
Naltrexone
Metabolite ID using PIF
Venlafaxine

Venlafaxine m/z 278

LC-MS of 5 µM Venlafaxine + Metabolite Standards
Metabolite ID using PIF
Venlafaxine

Parent drug
Venlafaxine
m/z 278

Metabolite I
m/z 264

Metabolite II
m/z 264

Metabolite III
m/z 250

Matching product ion spectra

Patent Pending

HighChem
<table>
<thead>
<tr>
<th>Precursor Ion Fingerprinting (PIF)</th>
<th>Venlafaxine</th>
<th>Metabolite I</th>
<th>Metabolite II</th>
<th>Metabolite III</th>
<th>Metabolite IV</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Venlafaxine" /></td>
<td><img src="image2" alt="Metabolite I" /></td>
<td><img src="image3" alt="Metabolite II" /></td>
<td><img src="image4" alt="Metabolite III" /></td>
<td><img src="image5" alt="Metabolite IV" /></td>
<td></td>
</tr>
<tr>
<td><img src="image6" alt="m/z 201" /></td>
<td><img src="image7" alt="m/z 201" /></td>
<td><img src="image8" alt="m/z 201" /></td>
<td><img src="image9" alt="m/z 201" /></td>
<td><img src="image10" alt="m/z 201" /></td>
<td></td>
</tr>
<tr>
<td><img src="image11" alt="m/z 215" /></td>
<td><img src="image12" alt="m/z 215" /></td>
<td><img src="image13" alt="m/z 215" /></td>
<td><img src="image14" alt="m/z 215" /></td>
<td><img src="image15" alt="m/z 215" /></td>
<td></td>
</tr>
<tr>
<td><img src="image16" alt="m/z 246" /></td>
<td><img src="image17" alt="m/z 246" /></td>
<td><img src="image18" alt="m/z 246" /></td>
<td><img src="image19" alt="m/z 246" /></td>
<td><img src="image20" alt="m/z 246" /></td>
<td></td>
</tr>
</tbody>
</table>

**Legend:**
- **m/z 201**
- **m/z 215**
- **m/z 246**

**MS^n**
- MS^3
- MS^4
- MS^5
Metabolite ID using PIF
Venlafaxine

RT: 0.00 - 16.98

NL: 2.33E7 m/z= 439.50-440.50 F: + c ESI Full ms [100.00-1000.00] MS 30Oct06_02

NL: 1.49E8 m/z= 249.50-250.50 F: + c ESI Full ms [100.00-1000.00] MS 30Oct06_02

NL: 2.00E8 m/z= 263.50-264.50 F: + c ESI Full ms [100.00-1000.00] MS 30Oct06_02

NL: 9.33E7 m/z= 277.50-278.50 F: + c ESI Full ms [100.00-1000.00] MS 30Oct06_02
Flavonoid Metabolite Identification by PIF

Plant extract
Satsuma Orange (*Citrus reticulata*)

ESI (+), Finnigan LCQ

PIF Library

Rutin
Gelsemine Metabolite Identification by PIF

Plant extract – Full scan
Yellow jasmine
(*Gelsemium sempervirens*)

Unknown

Unknown

Unknown

Unknown

ESI (+), LTQ Orbitrap

MS\(^1\)

MS\(^2\)

MS\(^3\)

MS\(^4\)
Determination of Molecular Structures Using Tandem Mass Spectrometry

Inventor: Robert Mistrik, Bratislava (SK)

Assignee: Highchem, Ltd., Bratislava (SK)

Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 27 days.

Appl. No.: 10/907,018
Filed: Oct. 14, 2004

Prior Publication Data
US 2006/0085142 A1 Apr. 20, 2006

Int. Cl.
G01N 3/00 (2006.01)
G01F 19/00 (2006.01)

Field of Classification Search
702/22, 702/27, 702/28, 250/281, 282, 292, 435/6, 435/7.1

See application file for complete search history.

References Cited

U.S. PATENT DOCUMENTS
5,072,115 A 12/1991 Zhou 250/281
6,455,321 B1 9/2002 Chase 436/173
6,623,935 B2 9/2003 Overney et al. 435/7.1
6,624,038 B1 9/2003 Raffa 250/282
6,674,069 B1 1/2004 Martin et al. 250/287
6,675,104 B2 1/2004 Panbe et al. 250/282

FOREIGN PATENT DOCUMENTS
WO WO 02/08600 12/2001

OTHER PUBLICATIONS

* cited by examiner

Primary Examiner—Bryan Bui

Attorney, Agent, or Firm—Weingarten, Schargin, Gagnebin & Lebovic LLP

Abstract

Methods for determining the chemical structures, substructures and/or ionic structural arrangements of unknown or partially structurally characterized compounds from mass spectrometric data are disclosed. The methods of the invention are directed to elucidating the molecular structures of compounds by identifying the structures of ions generated from those compounds through the use of tandem mass spectrometry and subsequently conducting a library search of structurally characterized tandem product spectra from known ions for specific matches. The collective structural information and/or alternative structural candidates are then compared to the determined ion structures and other structural characteristics using a structure generator and/or a structure assembly and/or a structure reduction system. The structure or substructure determination process can be continuously improved by adding additional data, e.g., information obtained from molecules previously structurally characterized using this method.

15 Claims, 3 Drawing Sheets
Determination of Ion Structures in Structurally Related Compounds Using Precursor Ion Fingerprinting

Michelle T. Sheldon, Robert Mistrik, Timothy R. Croley

Mass spectrometry (MS) is a well-established and powerful analytical technique. Analytical laboratories throughout the world rely on MS to identify unknown compounds, quantify ions, and determine pertinent structural information. Currently gas chromatography/mass spectrometry (GC/MS) is the most commonly used method for identification of unknown compounds [1]. GC/MS employs bank or transfer techniques (e.g., electron ionization (EI)), which are independent of model or manufacturer, largely reproducible from instrument to instrument, and consequently incorporate standard mass spectral libraries to identify and/or identify compounds [2].

A major advantage of GC/MS instruments available on the market and a comparable number of commercial spectral libraries is their ease of use with minimal spectral libraries needed to be analyzed by GC/MS [4]. In contrast, when a large number of samples, this can be significantly time-consuming and labor-intensive. More importantly, this means that while GC/MS spectral libraries are useful in solving problems, they are only applicable to a limited number of compounds [5].

Recently, liquid chromatography/mass spectrometry (LC/MS) employing electrospray ionization (ESI) has gained more use as an effective analytical tool. By comparison, ESI/MS can be used to analyze more classes of compounds (to include polar compounds), requires less sample preparation (as a derivatization step before analysis), is more sensitive, demonstrates increased specificity, and is quantitatively more accurate [2, 3, 4]. In addition, smaller induced dissociation (CID) or tandem mass spectrometry (MS/MS) techniques can be incorporated to provide fragmentation to elucidate considerable structural information pertinent to a compound [5]. Despite these advantages, neither has an untested, searchable fragmentation spectral library that is not confined to a certain type of instrumentation and/or specific class of compounds [7].

Activating spray-generated ESI/MS data to a database in a spectral library from different instrumental platforms is not easy. Many have made attempts to utilize already commercially available resources, while even more have tried to develop their own ESI/MS spectral libraries. ESI databases, such as the National Institute of Standards and Technology (NIST) and Wiley libraries, are fully evolved and contain an abundant number of compounds. In terms of reliability, this would seem like the most reasonable place to start. ESI spectra, however, in most cases, are very different than those of EI and full scan of a useful database [7].

In this study, we attempted to utilize only the NIST algorithm to search through their own in-house ESI/MS/MS spectra [9]. The software used for data acquisition in this study could not accept the precursor ion along with the product ions into the NIST database [9]. As a result, only the product ion(s) was/were searchable. Many
Identical molecular mass (unit), virtually identical isotopic pattern

For Confident Compound Identification MS³ is Needed !!!

Thermo Electron Finnigan LCQ Deca ESI+
Identical Molecular Formula = Identical Exact Mass, Identical MS²

**Spectral Libraries - Specificity**

**AMIDEPHRINE**

![AMIDEPHRINE Structure]

**C₁₀H₁₆N₂O₃S**

**BIOTIN**

![BIOTIN Structure]

**C₁₀H₁₆N₂O₃S**

**HighChem**
The logical data structure that best reflects spectra dependencies is a tree.

Total composite spectra – lost of the most valuable information
Product Ion Spectra Representation
Acknowledgment

• Spectral Library
  Ernst Pittenauer
  Vienna University of Technology, Vienna, Austria

• Software
  Juraj Lutisan, Milos Suchy
  HighChem, Ltd., Bratislava, Slovakia

• Fragmentation Library
  Andrea Belicova, Slavka Bekesova, Inrig Fux
  HighChem, Ltd., Bratislava, Slovakia

• Thermo Fisher Corporation, San Jose, CA
  Yingying Huang, Karen Salomon, Caroline Ding, Rose Herbold