

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

**Robert Mistrík**

---

HighChem Ltd.  
Bratislava, Slovakia

[www.highchem.com](http://www.highchem.com)

CoSMoS Boston 2009

**HighChem**

## Overview

### **Traditional Techniques**

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

### **New Method**

- Precursor Ion Fingerprinting (PIF)

## Mass Spectrometry is indirect structure elucidation method

From mass to shape



From mass to arrangement

m/z

147.03

183.03

201.08

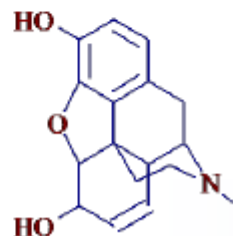
202.14

211.05

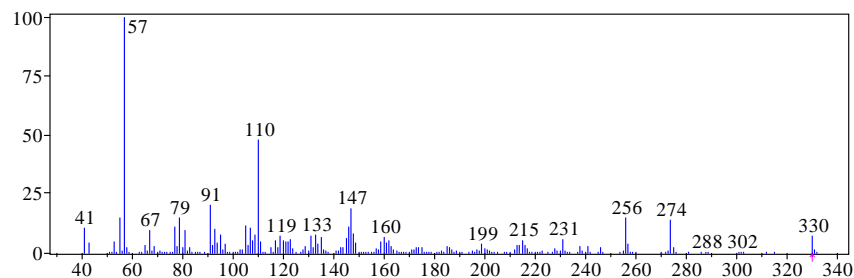
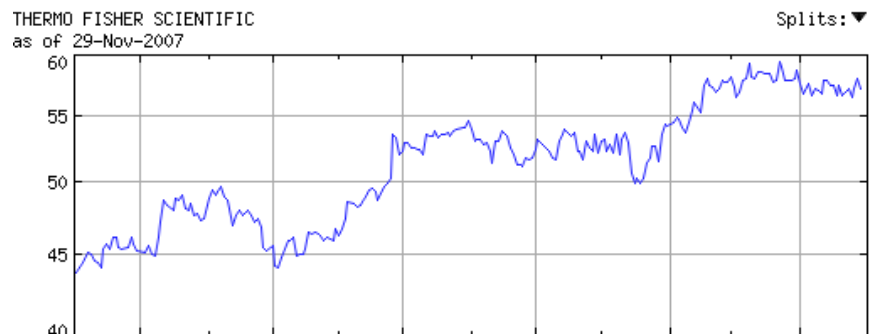
227.09

228.37

229.10



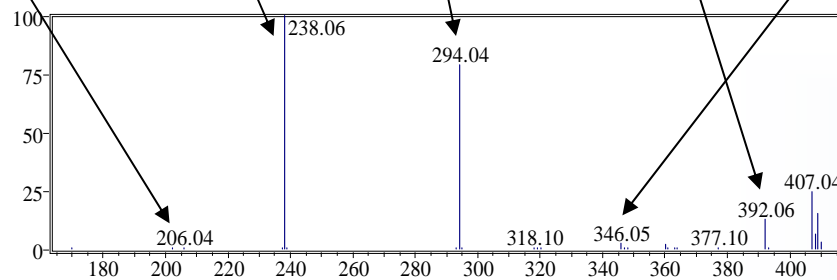
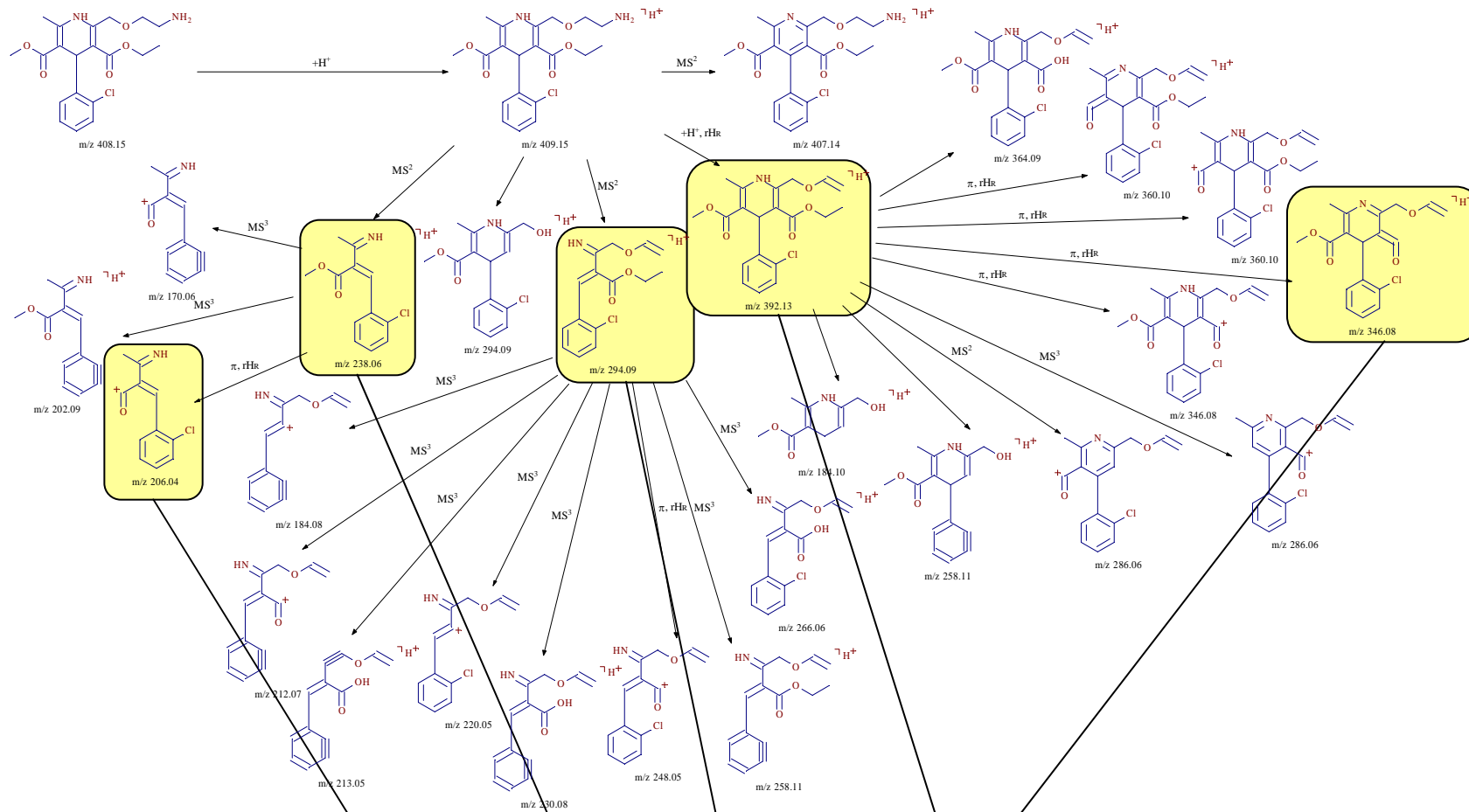
# Fragmentation Processes



“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”

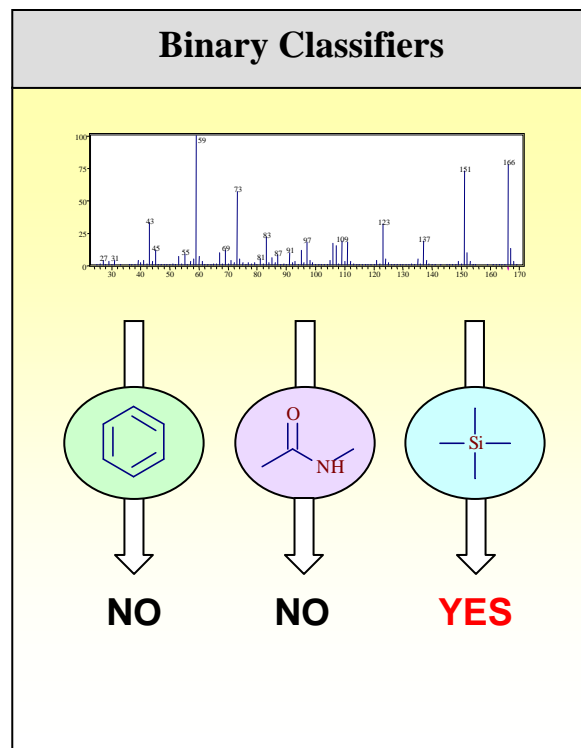
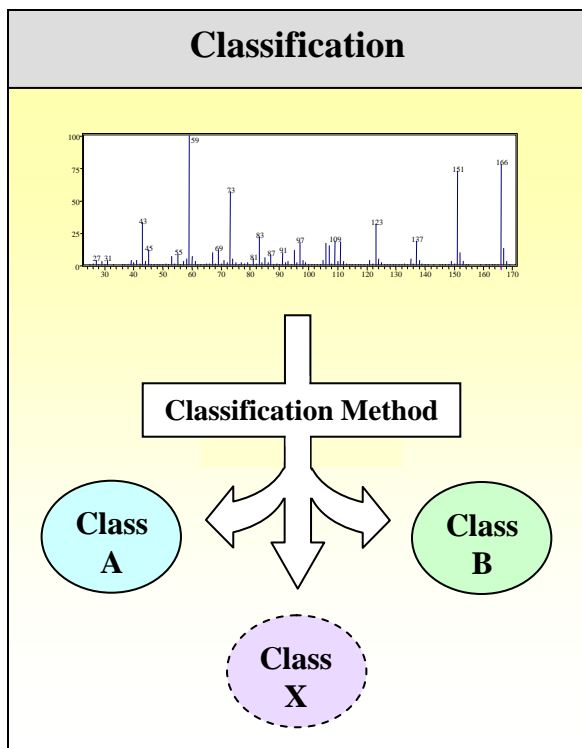
Klaus Biemann, Mass Spectrometry: Organic Chemical Applications, McGraw-Hill, New York, 1962, pp 72

# Fragmentation Processes



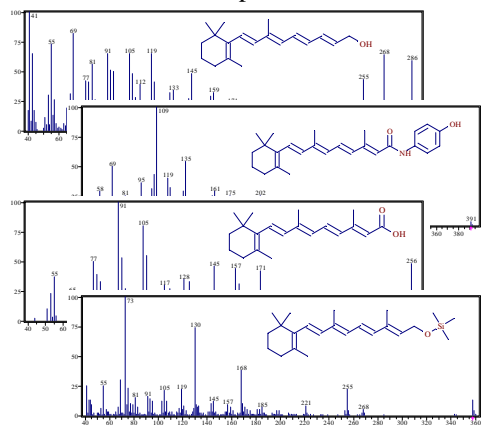
# Mass Spectra Classification

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.



# Mass Spectra Classification

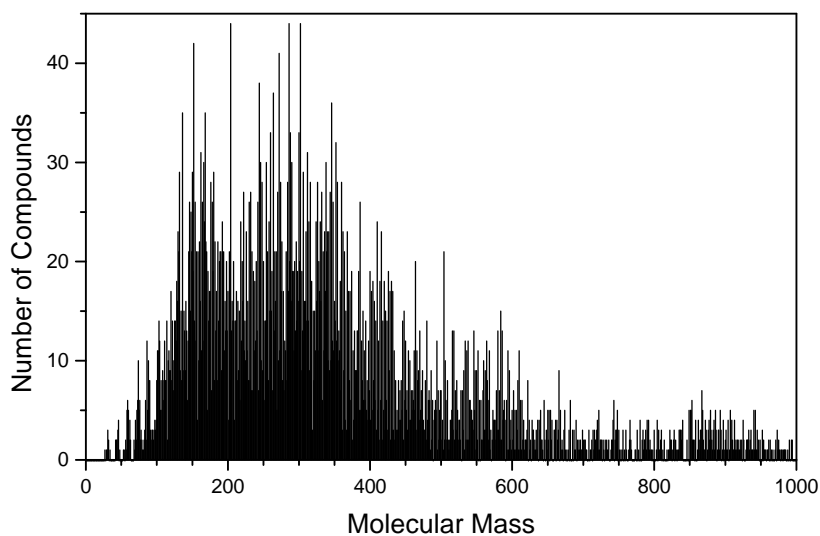
Vitamin A (Retinol) and its derivatives  
30 Spectra ◆



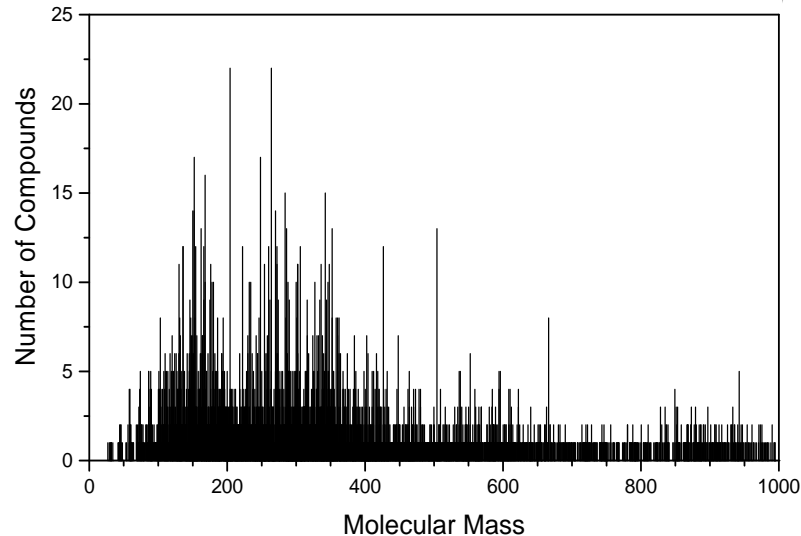
## Accurate mass search against structural libraries

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

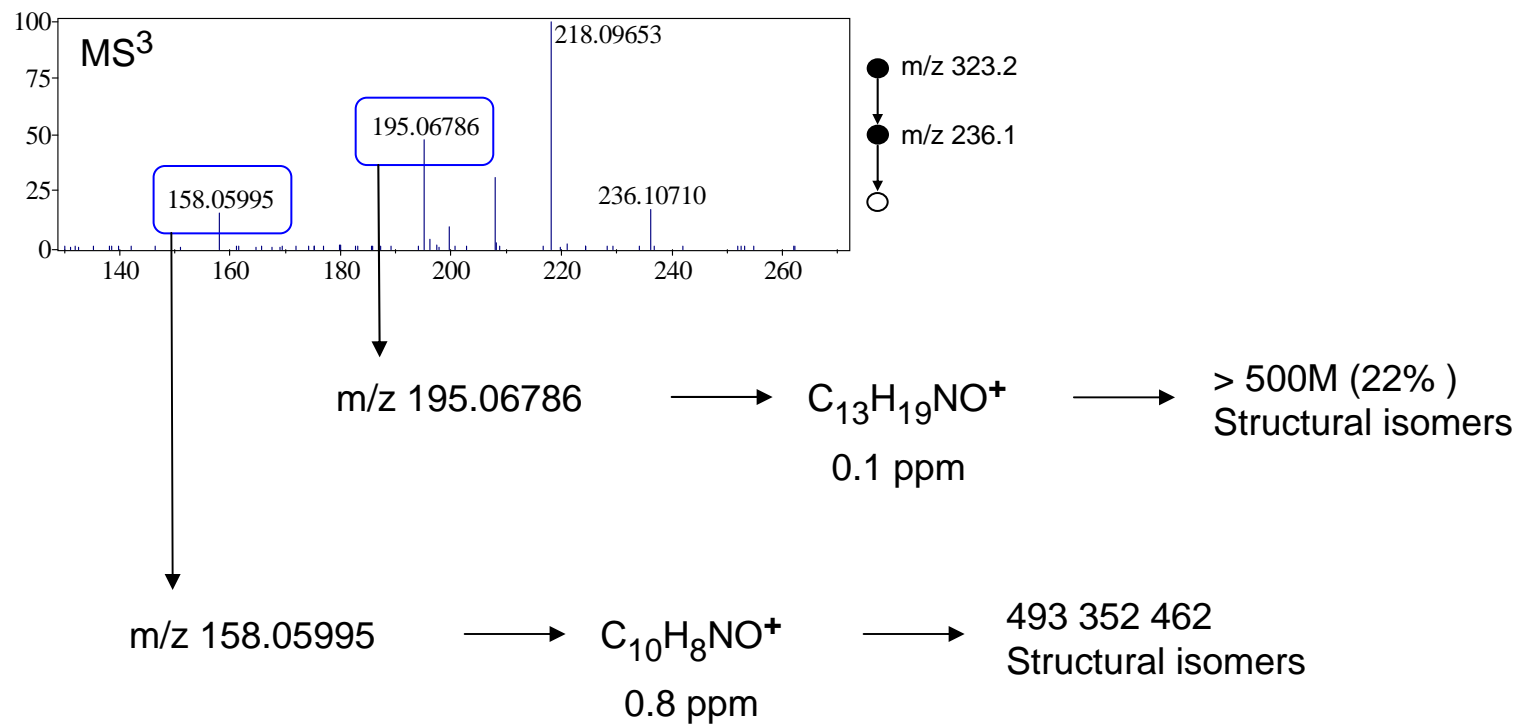
Unit resolution



$\Delta(m/z) = 0.00001$



## Immense Number of Structural Isomers

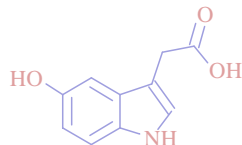
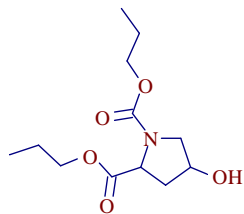
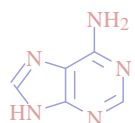


MS<sup>3</sup> Spectrum:  
Formula Generator:  
Structure Isomers:

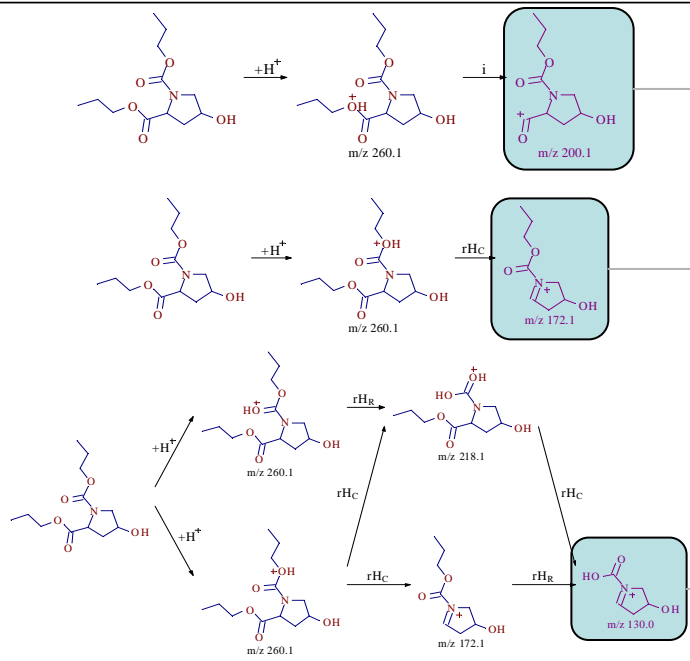
ESI (+), LTQ Orbitrap  
Mass Frontier 6.0  
Molgen 3.0

# Consistency checking between library structure and mass spectrum

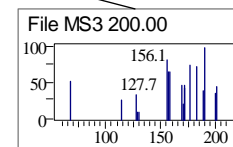
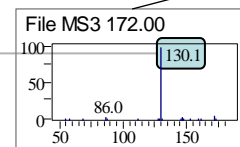
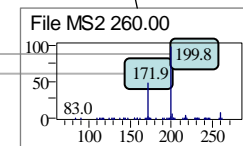
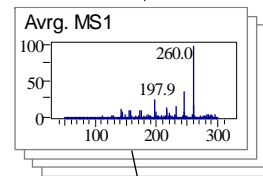
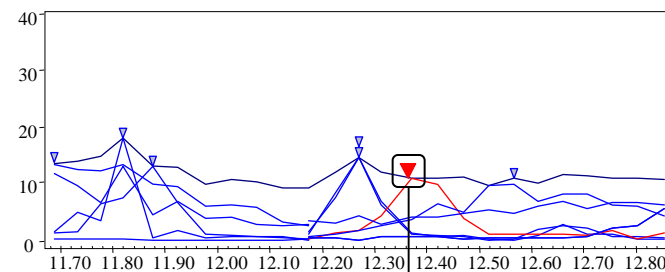
## Database of known human metabolites



## Automated of Prediction Fragments

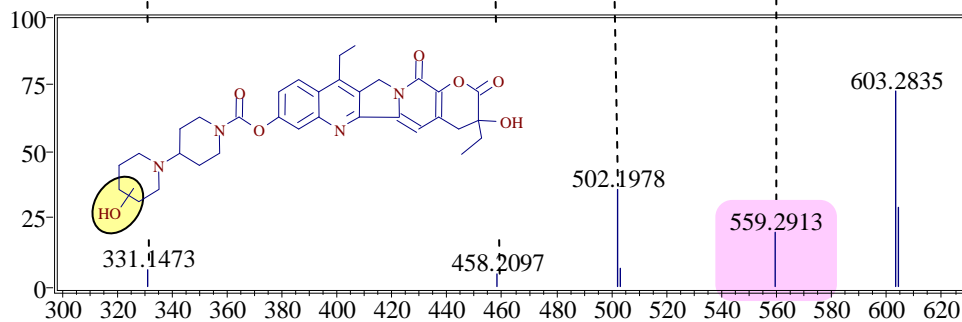
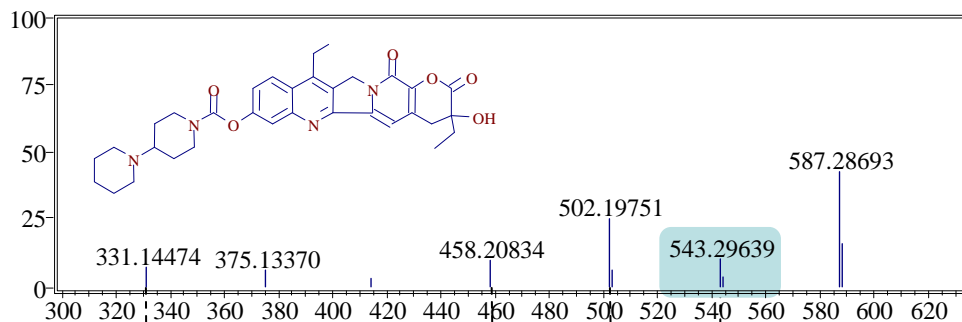


## Unknown Component in Biofluid

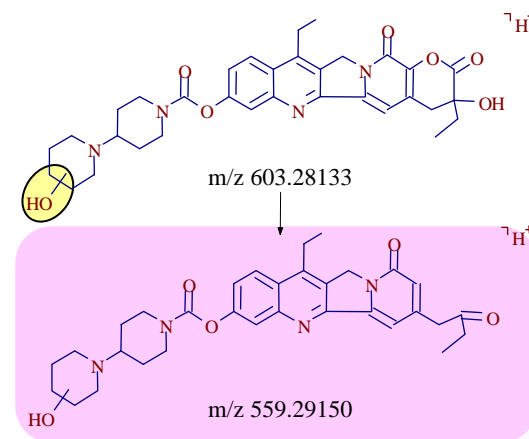
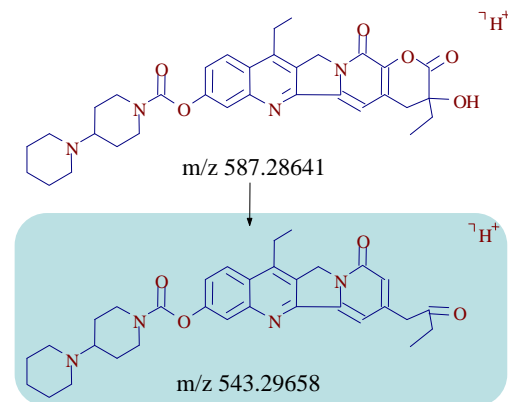


# Metabolite ID using Predictive Fragmentation

## MS<sup>2</sup> of Irinotecan (parent drug)



## MS<sup>2</sup> of Irinotecan metabolite (M6)

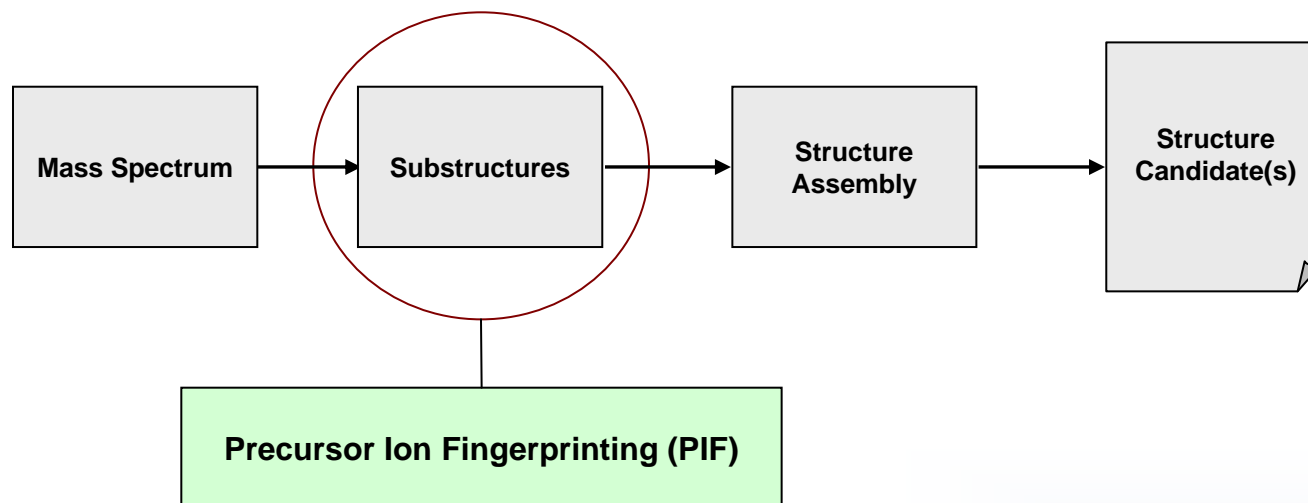


# Precursor Ion Fingerprinting (PIF)

## A new concept for the interpretation of mass spectra

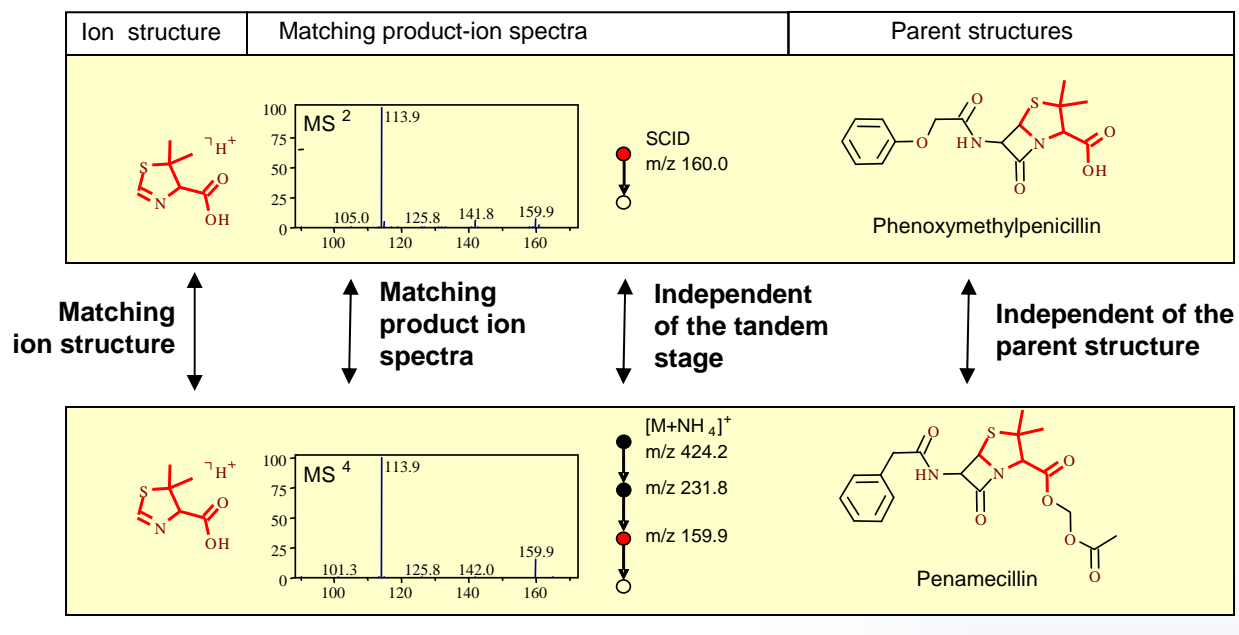
Efficient structure characterization of small molecules remains a huge bottleneck in an array of disciplines.

Principle of the structure elucidation using PIF



# Precursor Ion Fingerprinting (PIF) Theory

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.

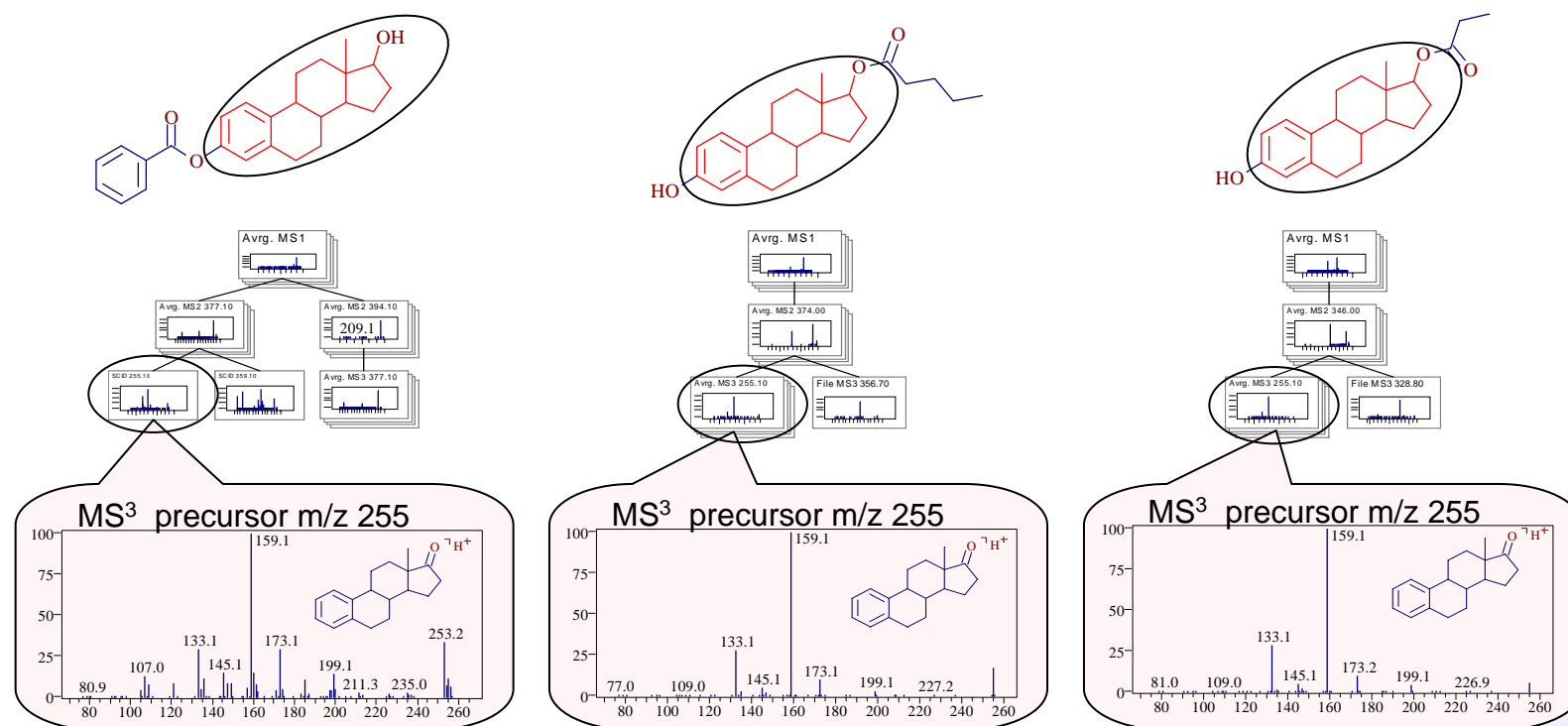


# Precursor Ion Fingerprinting (PIF)

## Quasi-equilibrium theory:

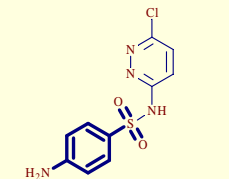
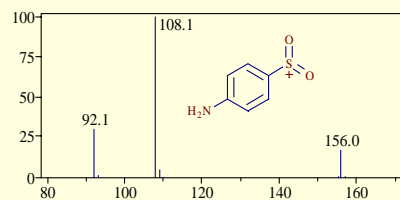
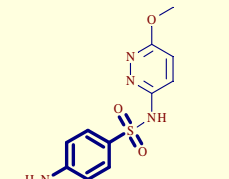
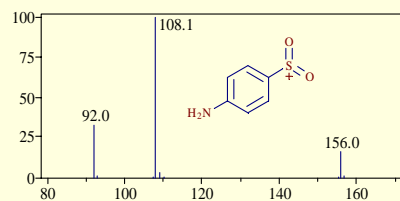
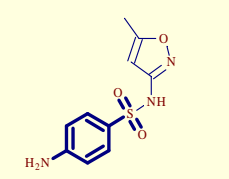
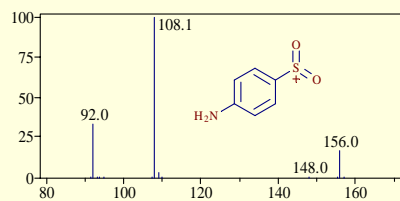
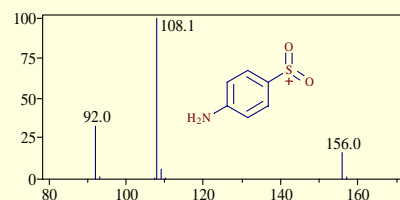
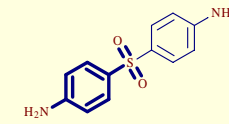
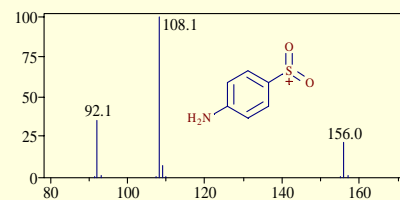
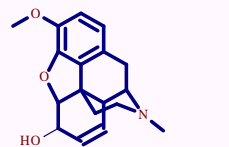
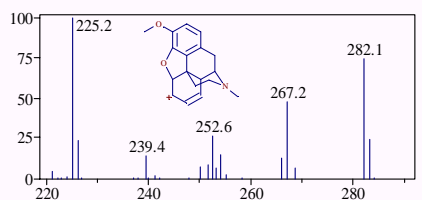
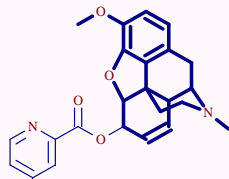
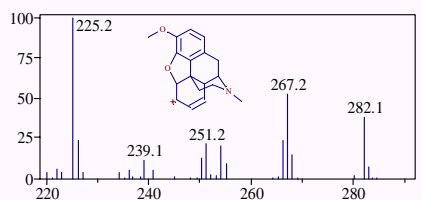
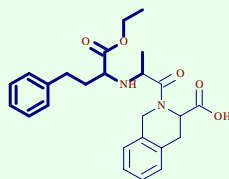
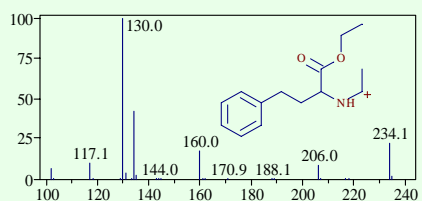
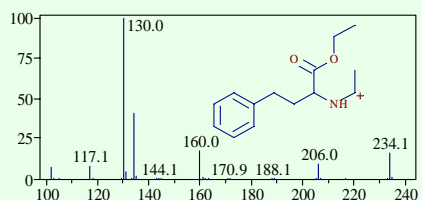
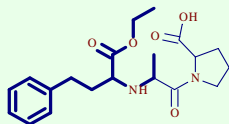
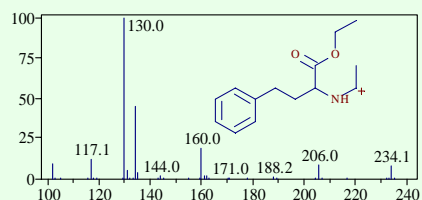
“The probabilities of the various possible decomposition products of an ion depend upon its structure, internal energy, and the energy deposited during ion activation, but are independent of the ionization method, the structure of the precursor ion, and the formation mechanism of the ion undergoing decomposition.”

F. McLafferty, F. Turecek, Interpretation of Mass Spectra, 1993, 4<sup>th</sup> Edition, USB



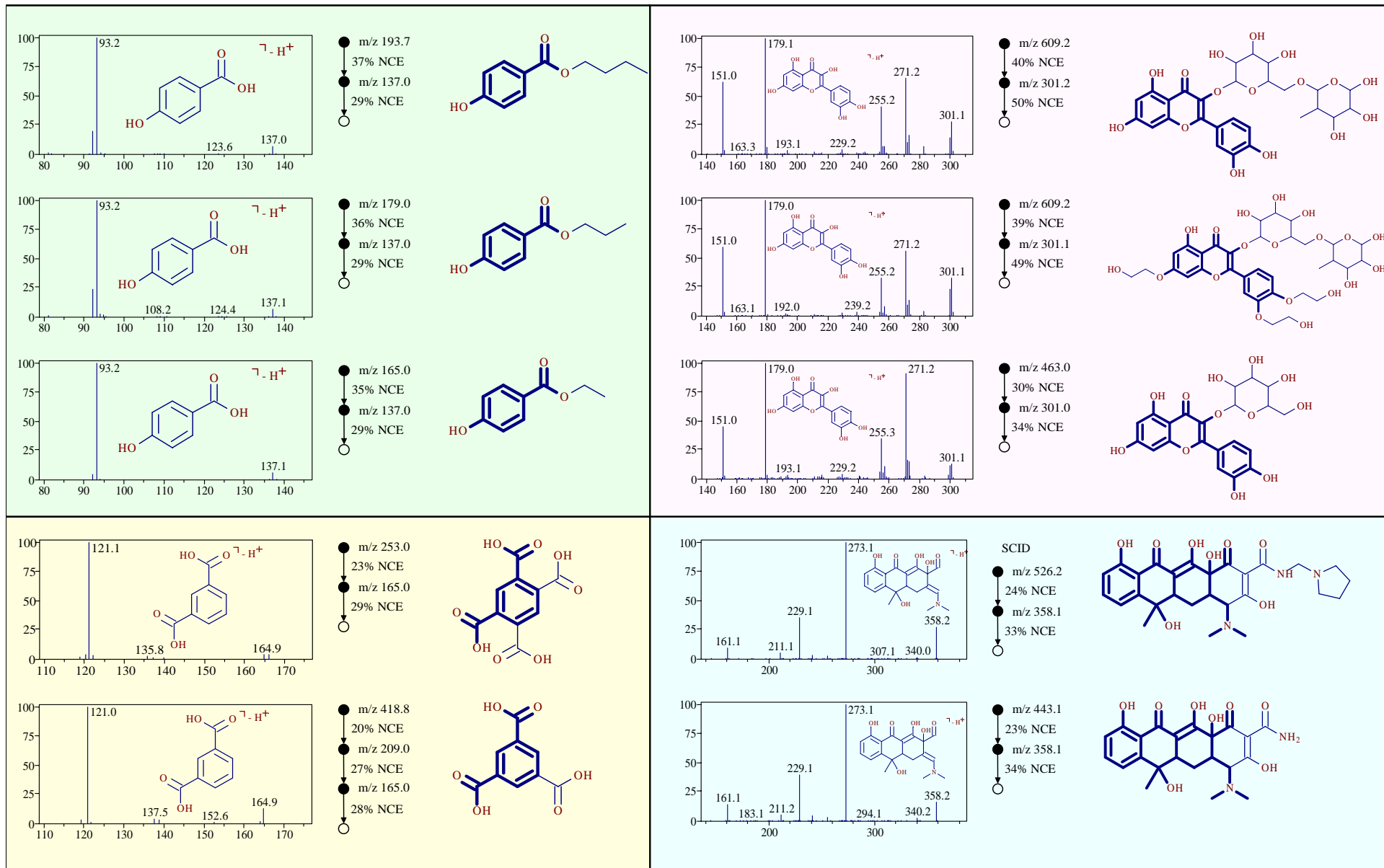
# Precursor Ion Fingerprinting (PIF)

ESI (+), Finnigan LCQ

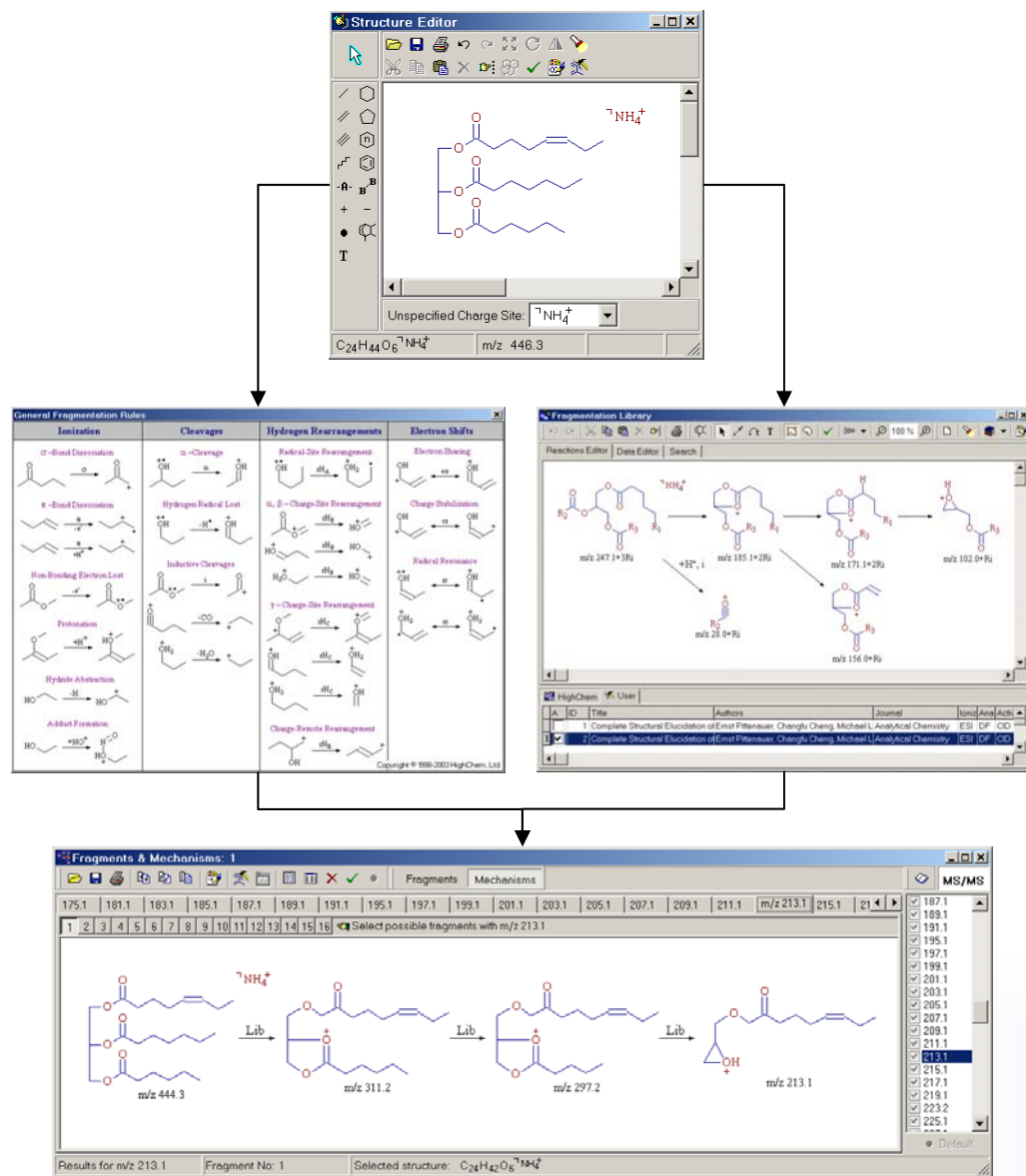


# Precursor Ion Fingerprinting (PIF)

ESI (-), Finnigan LCQ

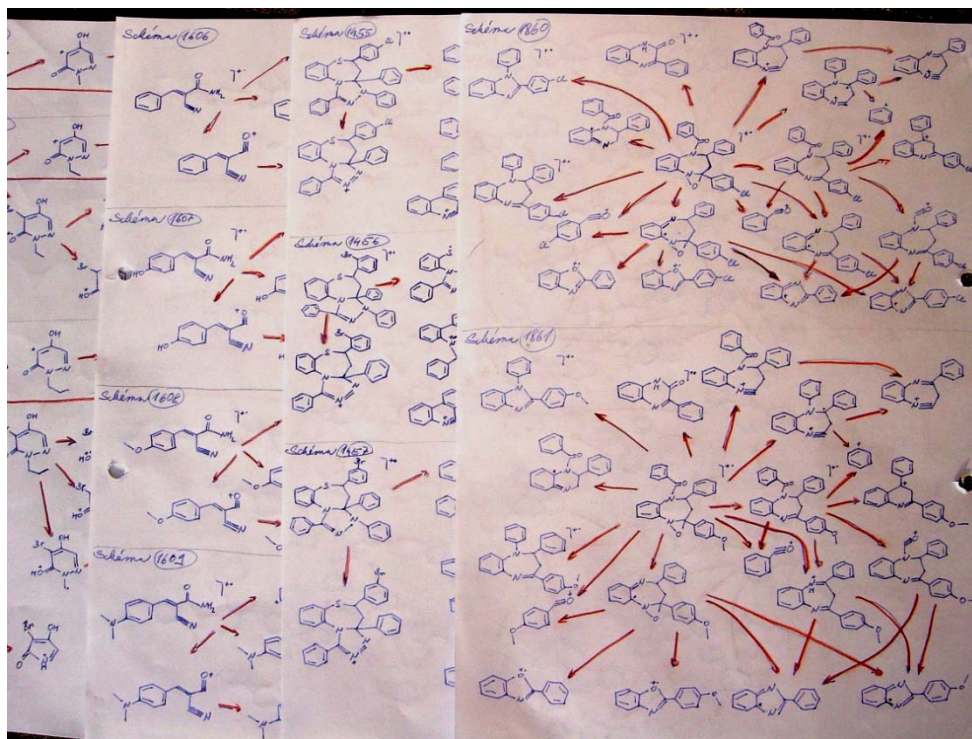


# Prediction of Fragmentation Pathways



## Mass Frontier Fragmentation Library™

Total number of	Mass Frontier 5.0	Mass Frontier 6.0
Fragmentation Schemes	20.003	30.936
Individual Reactions	87.966	129.229
Chemical Structures	101.850	151.762
Decoded Mechanisms	81.389	120.029



## Fragmentation Library™

	Source	Volume	Year
1.	JASMS (Journal of the American Society for Mass Spectrometry)	1-17	1990-2006
2.	IJMSIP (International Journal of Mass Spectrometry and Ion Physics)	1-53	1968-1983
	IJMSIP (International Journal of Mass Spectrometry and Ion Processes)	54-175	1983-1998
	IJMS (International Journal of Mass Spectrometry)	176-255	1998-2006
3.	RCM (Rapid Communications in Mass Spectrometry)	1-20	1987-2006
4.	JMS (Journal of Mass Spectrometry)	30-41	1995-2006
5.	OMS (Organic Mass Spectrometry)	1-29	1968-1994
6.	JMSSJ (Journal of the Mass Spectrometry Society of Japan)	11-27 29-30 37-48 50-53	1964-1979 1981-1982 1989-2000 2002-2005
7.	MSR (Mass Spectrometry Reviews)	1-25	1981-2006
8.	EJMSBMER (European Journal of Mass Spectrometry in Biochemical, Medicine, and Environmental. Research)	1-2	1980-1982
9.	BMS (Biomedical Mass Spectrometry)	1-12	1974-1985
	BEMS (Biomedical and Environmental Mass Spectrometry)	14-19	1987-1990
	BMS (Biological Mass Spectrometry)	20-23	1991-1994
10.	JC (Journal of Chromatography)	181-536	1980-1991
11.	EJMS (European Journal of Mass Spectrometry)	4	1998



# User provided fragmentation mechanism can improve fragment prediction for metabolites

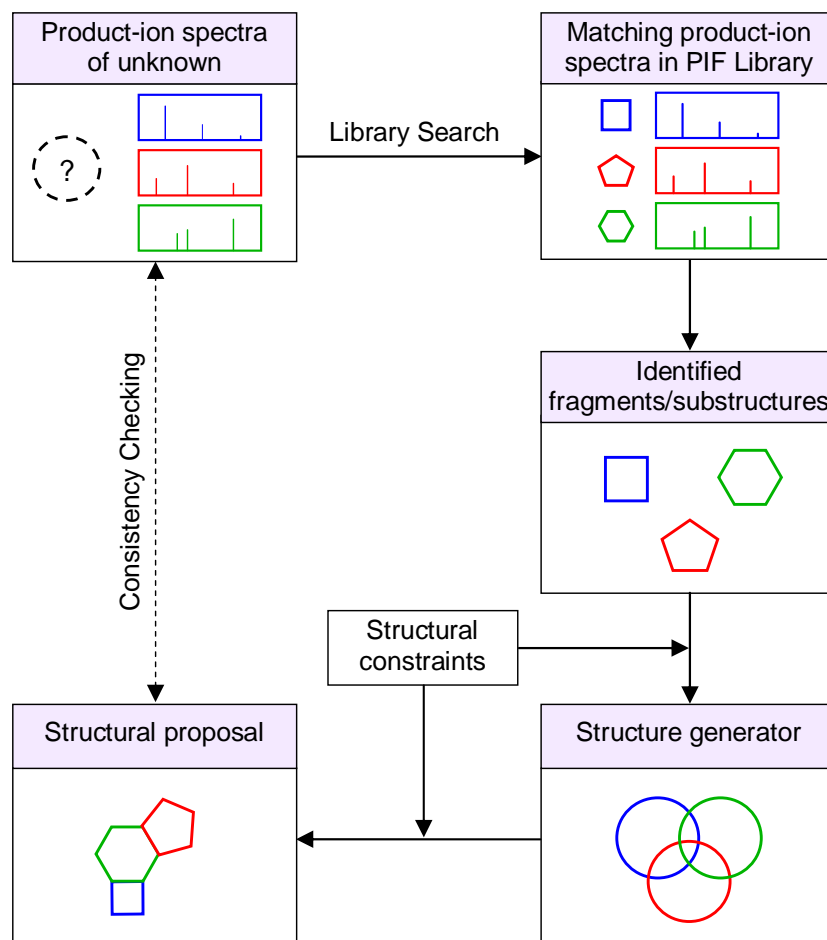
The screenshot displays the HighChem Mass Frontier 5.0 interface. The top window, 'Fragmentation Library', shows a chemical reaction where a parent drug (m/z 587) loses CO<sub>2</sub> to form a metabolite (m/z 543). The reaction is labeled with <sup>7</sup>H<sup>+</sup> on both sides. Below this, a table lists the parent drug as Irinotecan. The bottom window, 'Database Manager: 1', shows a mass spectrum with several peaks. A red box highlights a peak at m/z 586, which is identified as 'Metabolite M6 Irinotecan' in the 'Structures' table below. The table has columns for ID Num., Mol. Mass, Formula, and Name.

ID Num.	Mol. Mass	Formula	Name
1	586	C <sub>33</sub> H <sub>38</sub> N <sub>4</sub> O <sub>6</sub>	Metabolite M6 Irinotecan

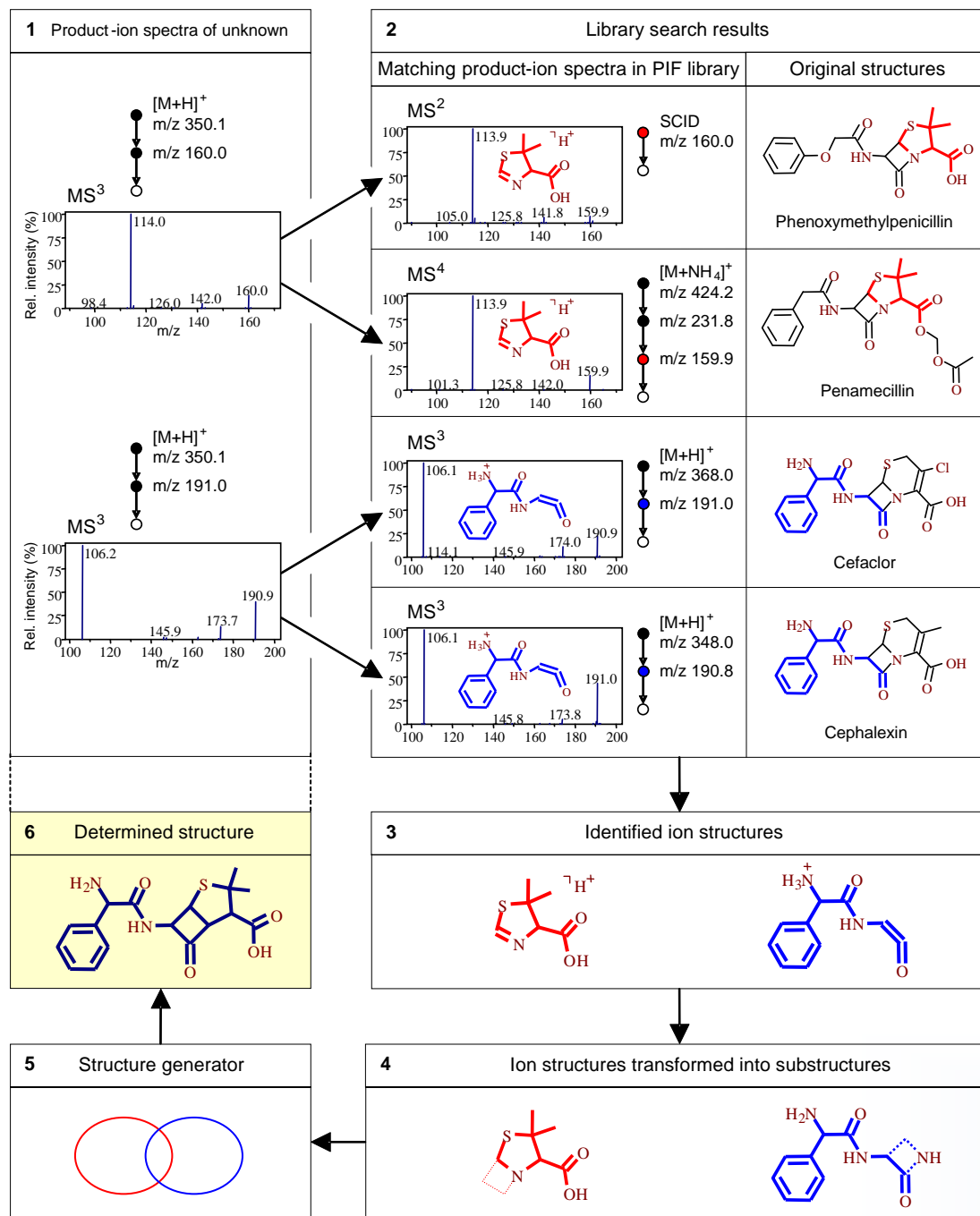
User suggested fragmentation mechanism for parent drug

User's mechanism applied to predicted metabolite structure

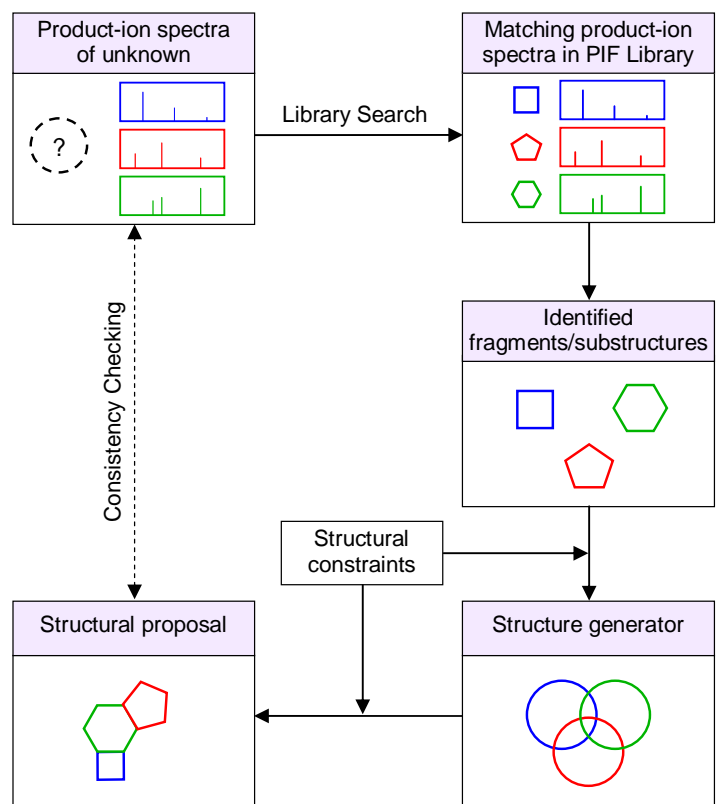
# Automated Interpretation of Mass Spectra



Patent pending



# Automated Interpretation of Mass Spectra

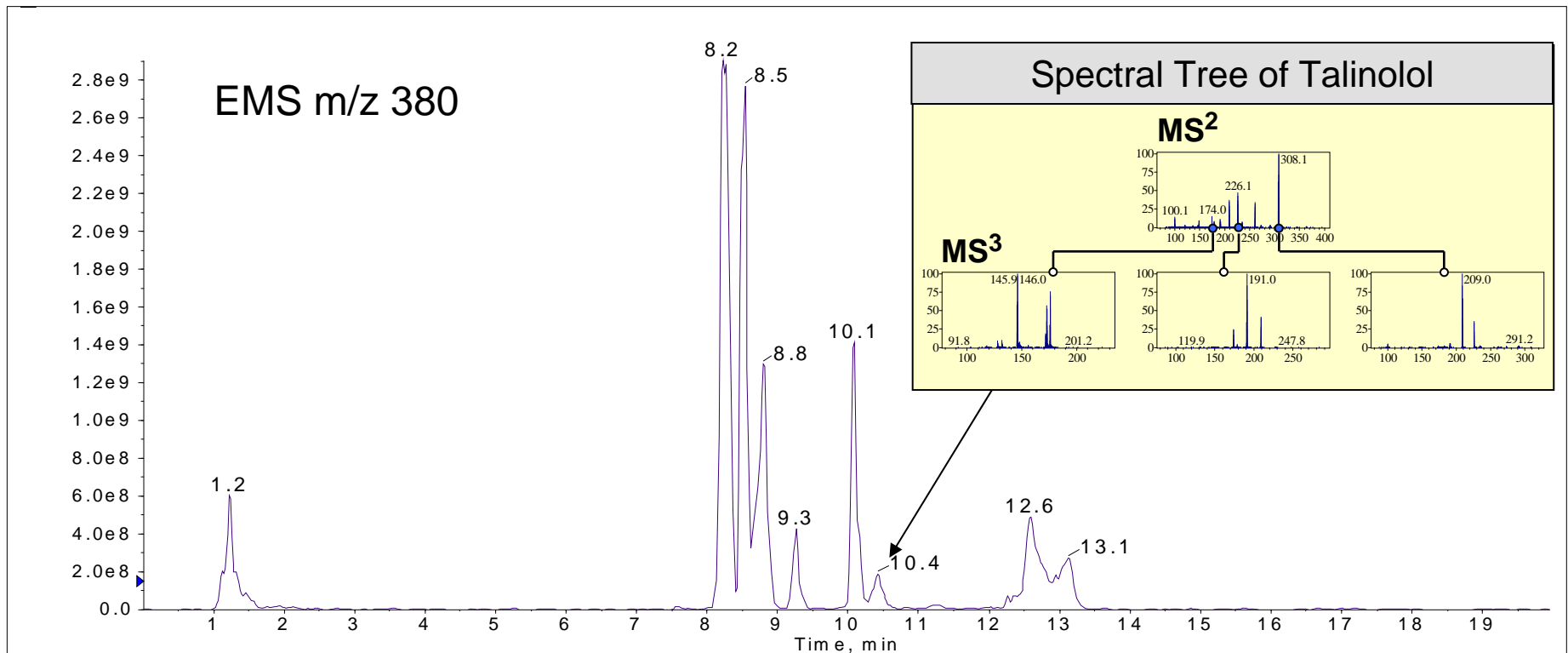


Patent pending

**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



# Fragment structure assignment

Parent drug  
(Library)

HighChem Mass Frontier 4.0

Fragmentation Library

Reactions Editor | Data Editor | Search

Talinolol  $m/z$  364.3

$m/z$  209.1

$m/z$  226.1

$m/z$  308.2

+H<sup>+</sup>, rHb

A	ID	Title	Authors	Journal	Ioniz. Method	Analyz.
	7					
	8					
9		Talinolol	Gérard Hopfgartner		ESI	IT

Fragments & Mechanisms: 2

209.1 | 226.1 | 291.2 |  $m/z$  308.2 | 380.3

1

$m/z$  380.3

$m/z$  308.2

Results for  $m/z$  308.2 | Fragment No: 1 | Selected structure: none

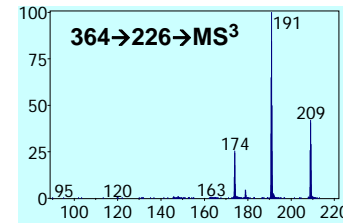
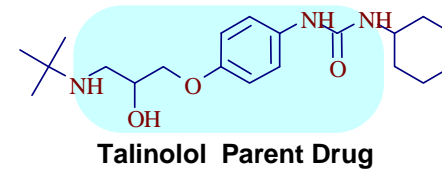
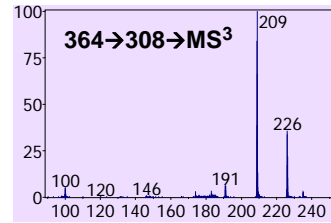
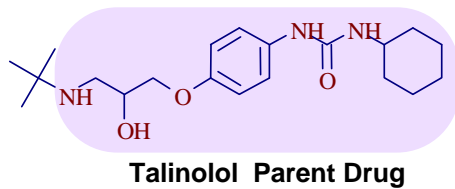
Fragmentation Library™:

- HighChem library 100,000 mechanisms
- User defined mechanism

Fragmentation reactions of parent drug stored in fragmentation library are applied to prediction of fragments for parent drug metabolites

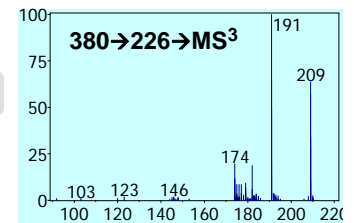
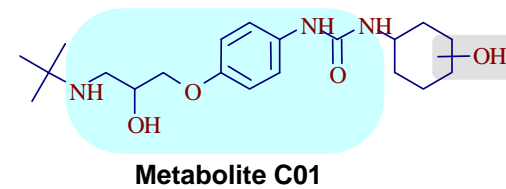
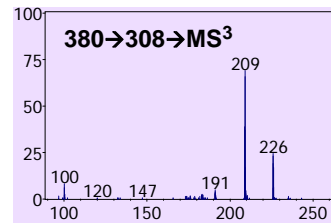
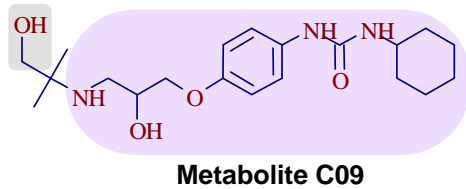
Metabolite  
(Prediction)

# Metabolite ID using PIF Talinolol



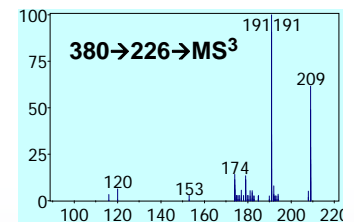
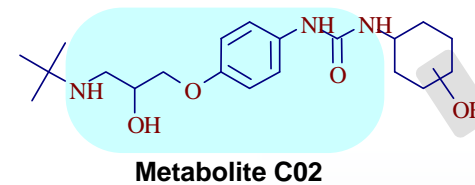
↕ Matching product ion spectra

↕ Matching product ion spectra

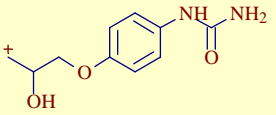
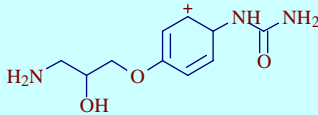
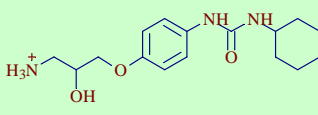
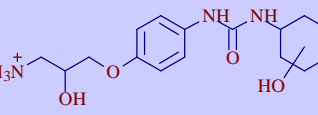
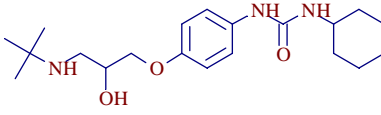
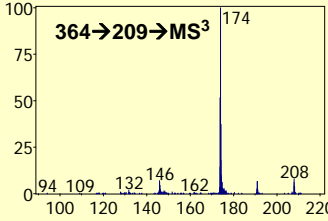
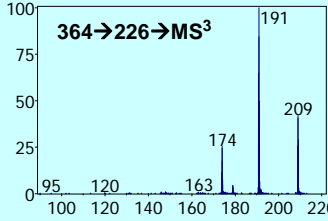
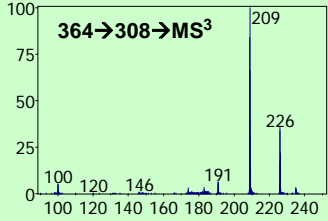
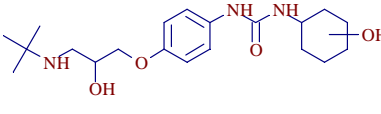
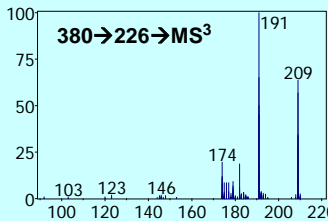
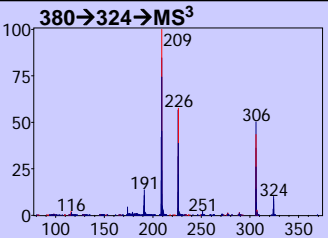
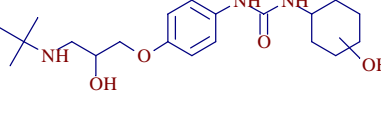
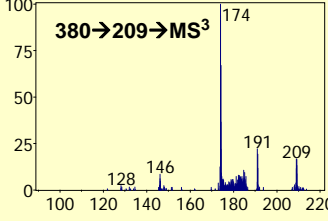
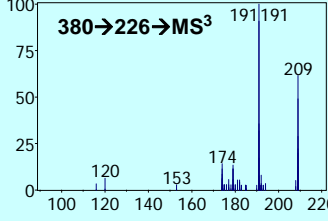
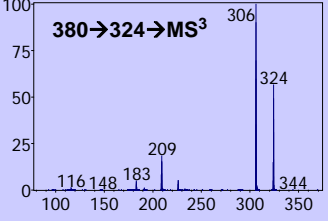
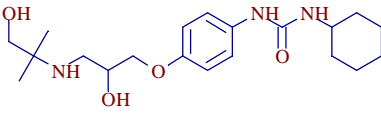
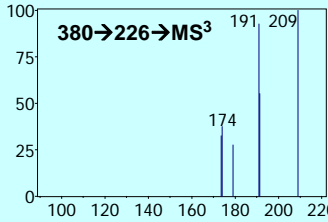
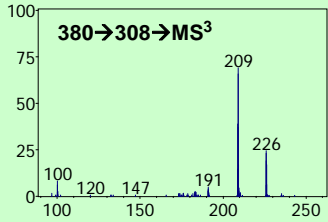


↕ Matching product ion spectra

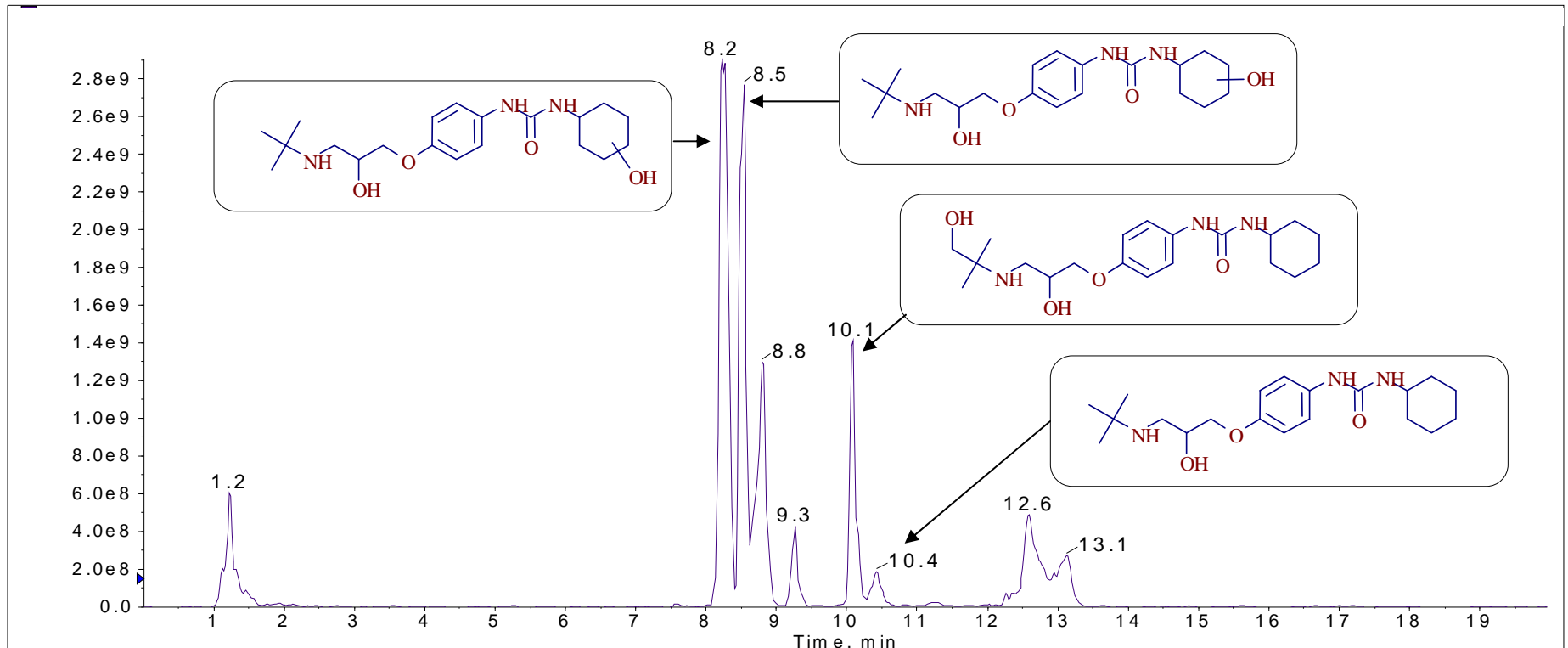
Patent Pending



Matching product-ion spectra (MS<sup>3</sup>) reveal common substructures of parent drug and metabolites

		m/z 209	m/z 226	m/z 308	m/z 324
	Metabolite Structure Assignment				
Talinolol					
Metabolite C01					
Metabolite C02					
Metabolite C09					<p>Intensity differences due to different hydroxyl group position</p>

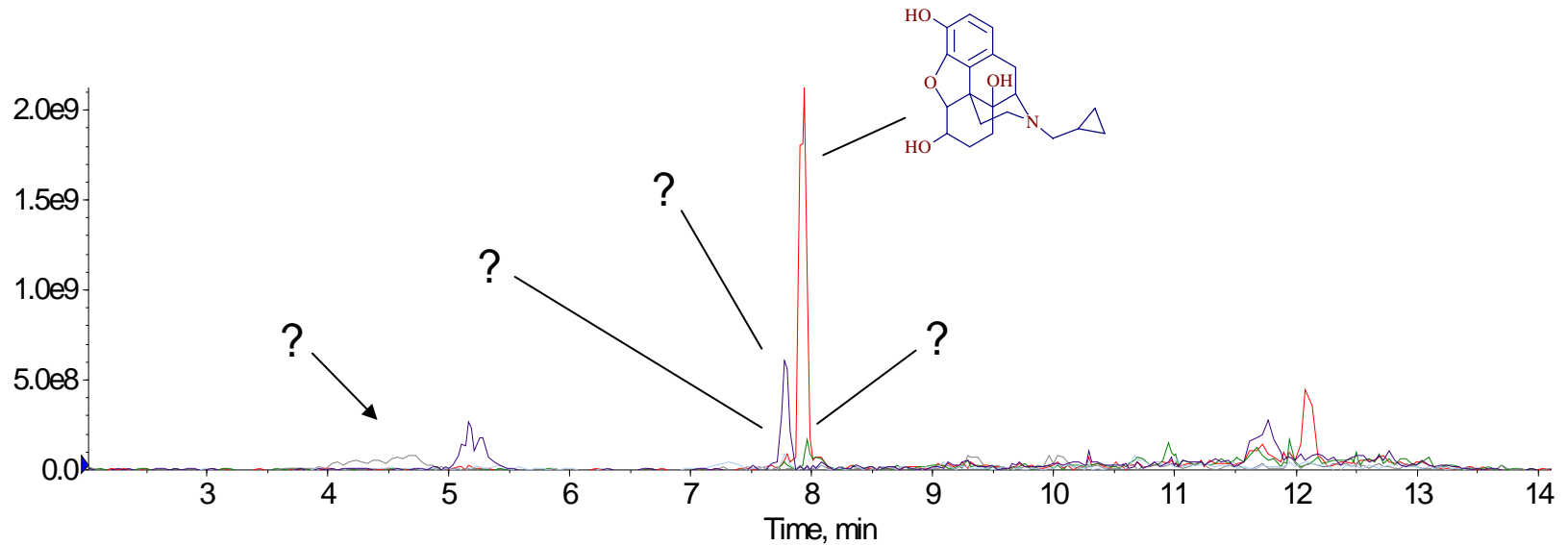
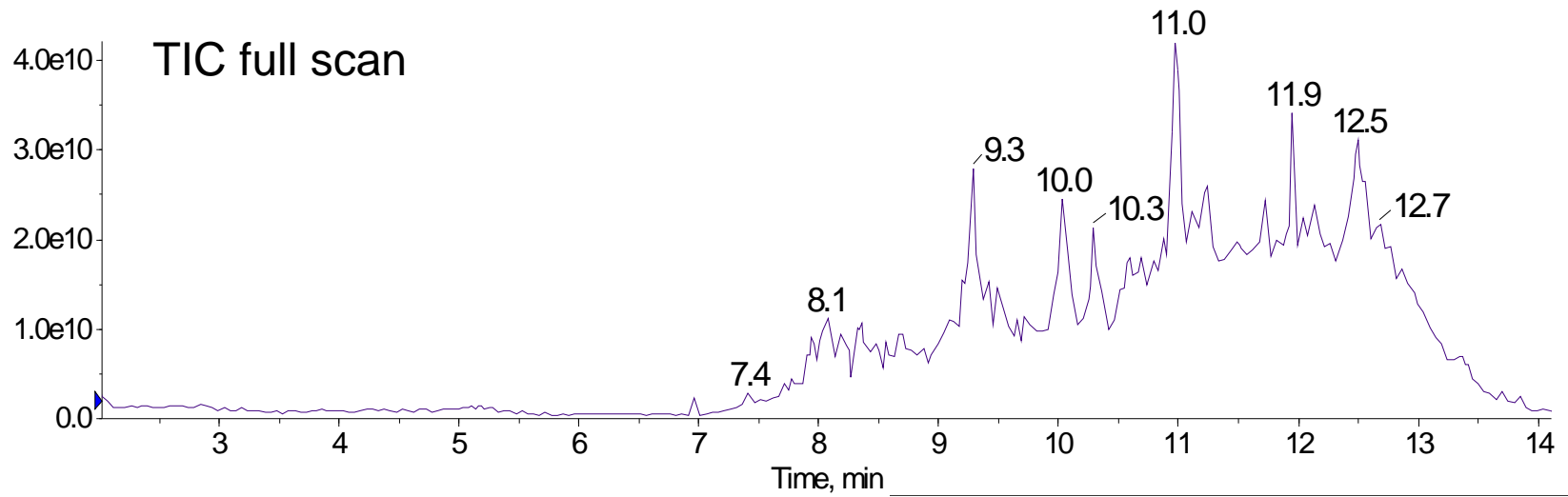
# Metabolite ID using PIF Talinolol



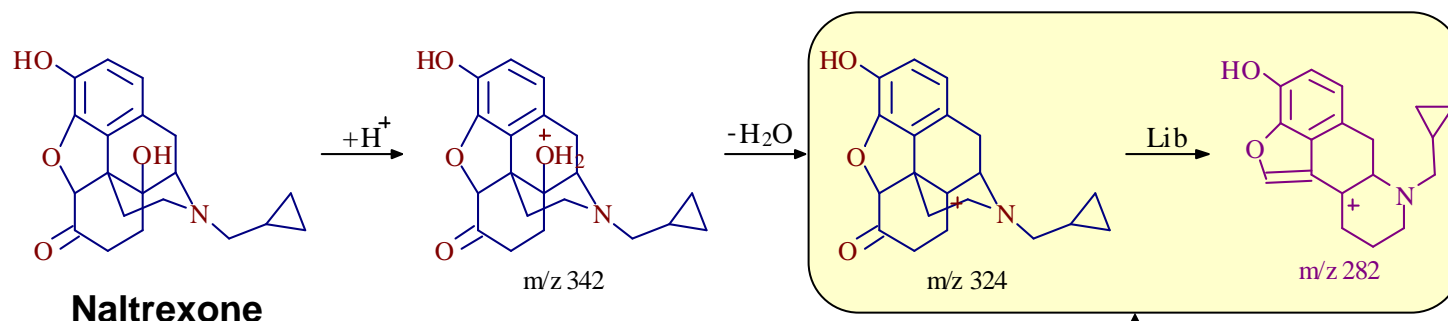
PIF concept allows high throughput, semi-automated qualitative analysis of structurally related compounds (xenobiotic or endogenous metabolites)

Data provided by G. Hopfgartner, University of Geneva, Switzerland

# Metabolite ID using PIF Naltrexone

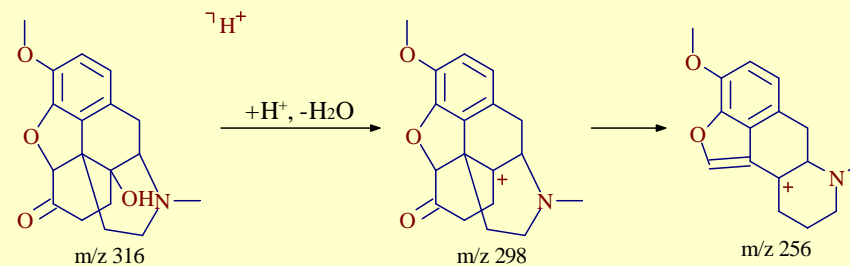


# Prediction of fragmentation pathways of Naltrexone using HighChem Mass Frontier and Fragmentation Library™



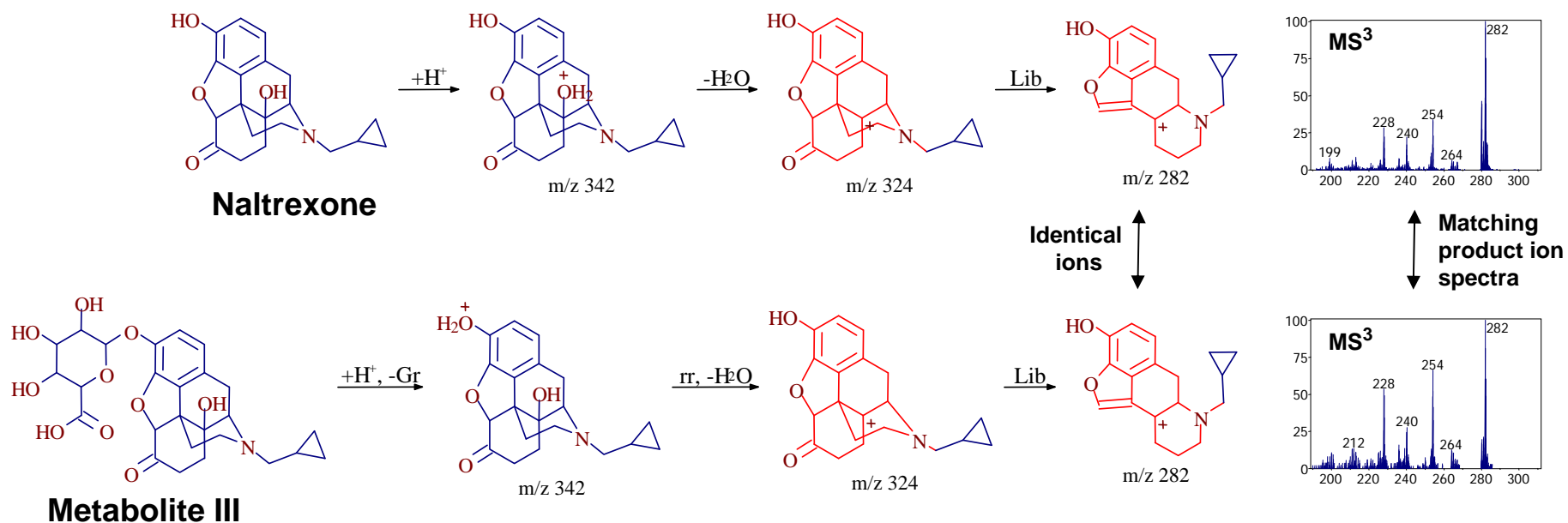
## HighChem Fragmentation Library™

ID: 6.735 of 89.326



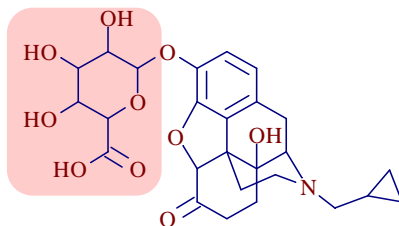
Pennanen, K.; Kotiaho, T.; Huikko, K.; Kostianen, R.;  
 Identification of ozone-oxidation products of oxycodone by  
 electrospray ion trap mass spectrometry. *Journal of Mass  
 Spectrometry* **2001**, 36, 795.

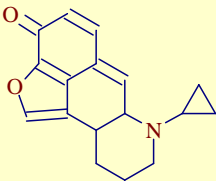

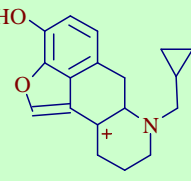
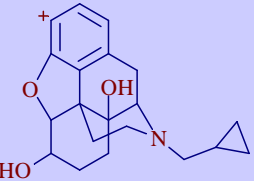
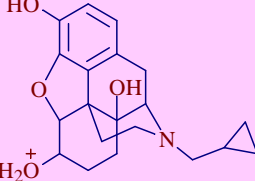
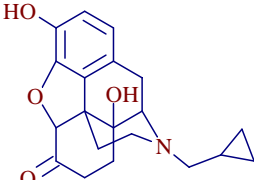
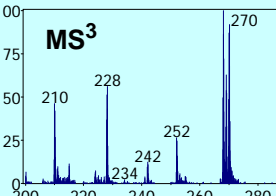
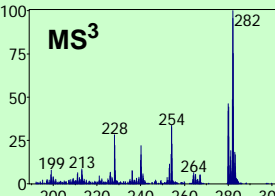
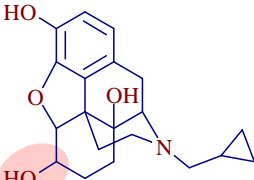
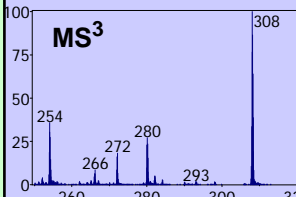
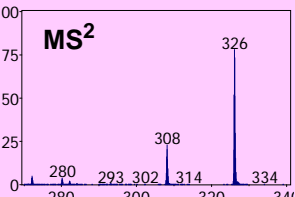
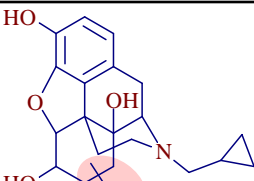
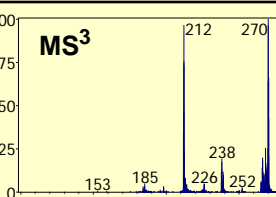
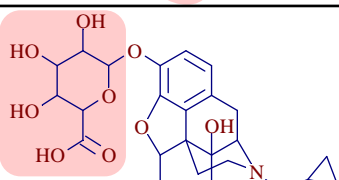
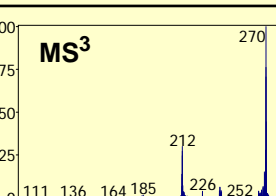
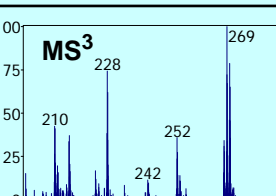
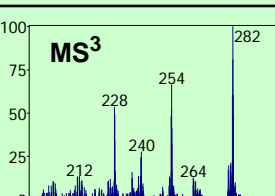
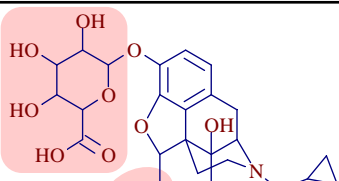
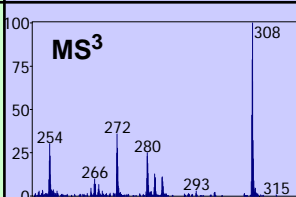
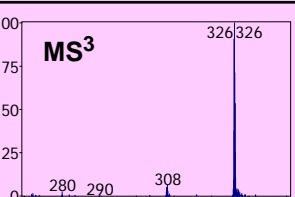
# Metabolite ID using PIF Naltrexone



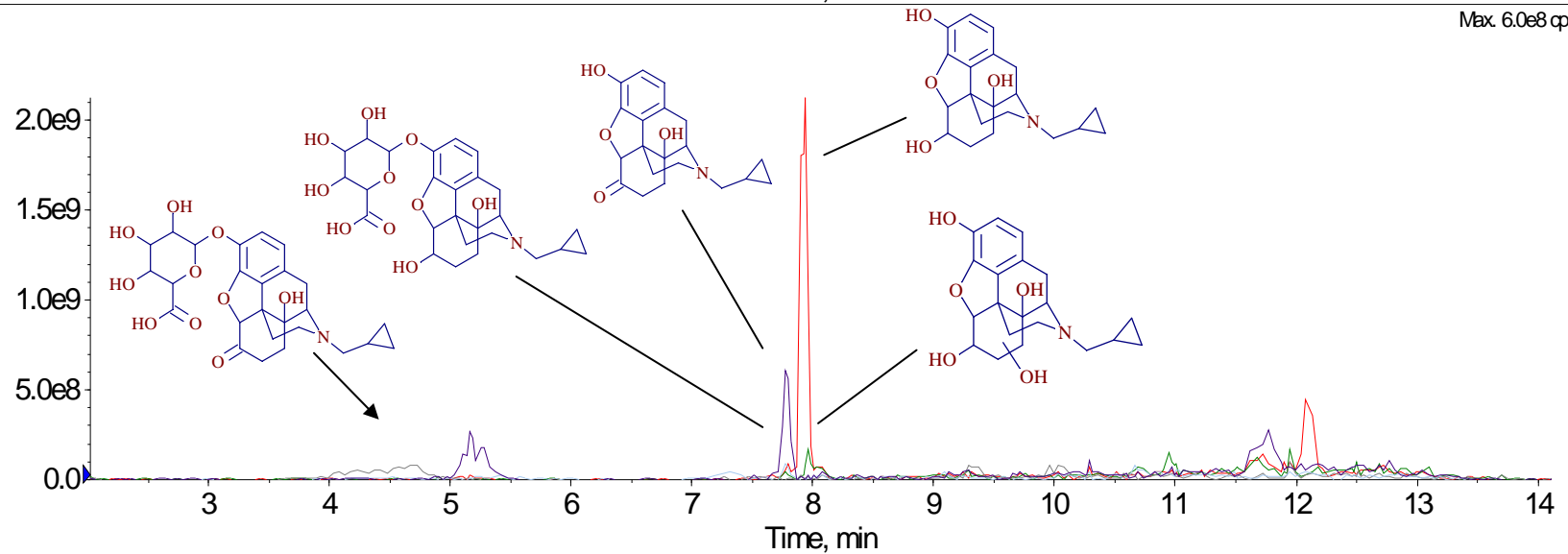
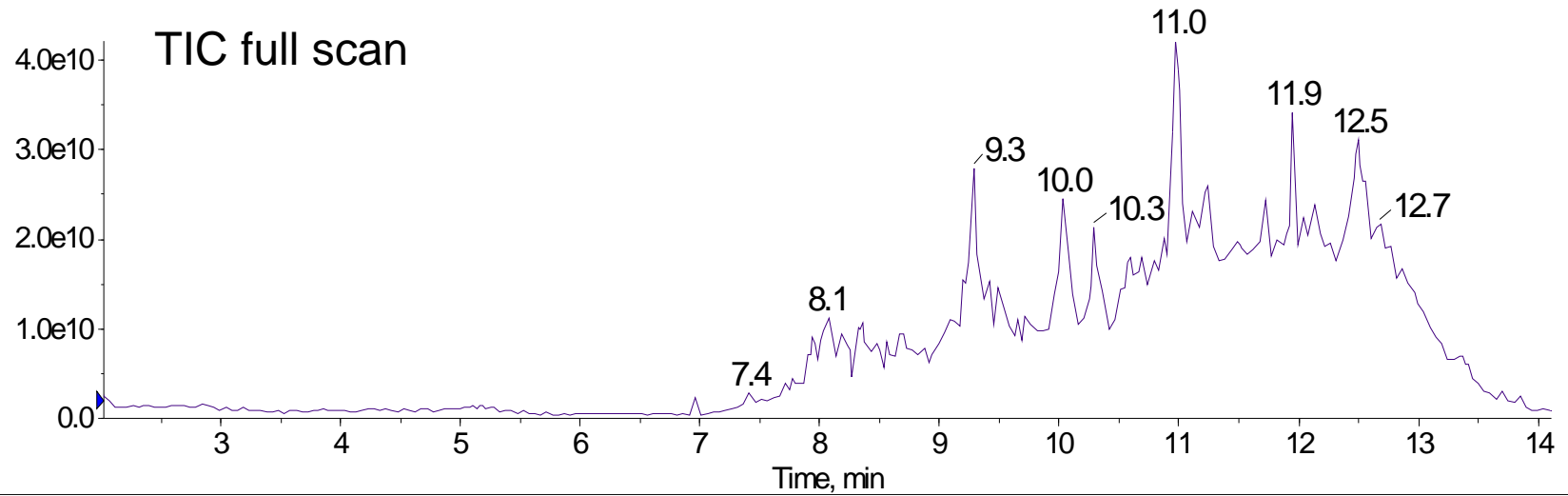
Structurally identical ions yield matching product ion spectra

Site of glucoronization can be determined using fragmentation prediction and PIF

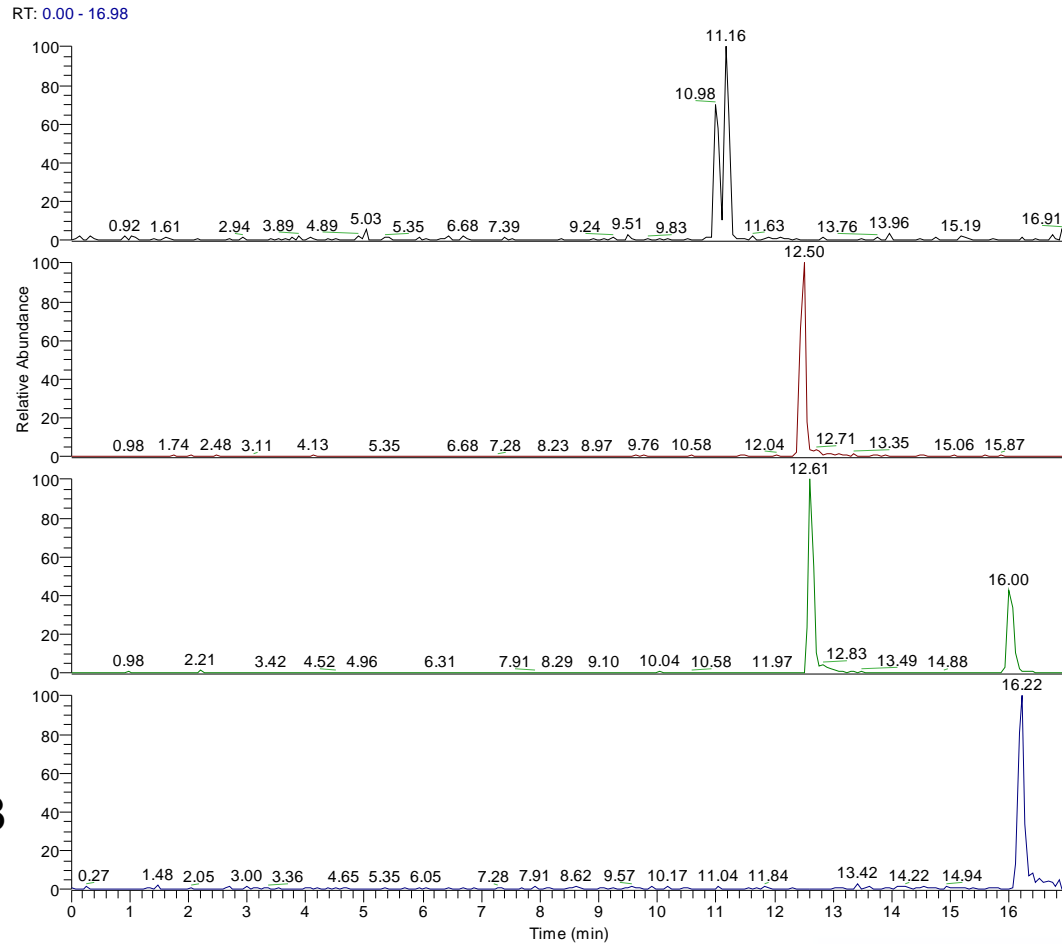


		m/z 267	m/z 270	m/z 282	m/z 326	m/z 344
	<b>Metabolite Structure Assignment</b>					
<b>Naltrexone</b>						
<b>Metabolite I</b>						
<b>Metabolite II</b>						
<b>Metabolite III</b>						
<b>Metabolite IV</b>						

# Metabolite ID using PIF Naltrexone



# Metabolite ID using PIF Venlafaxine

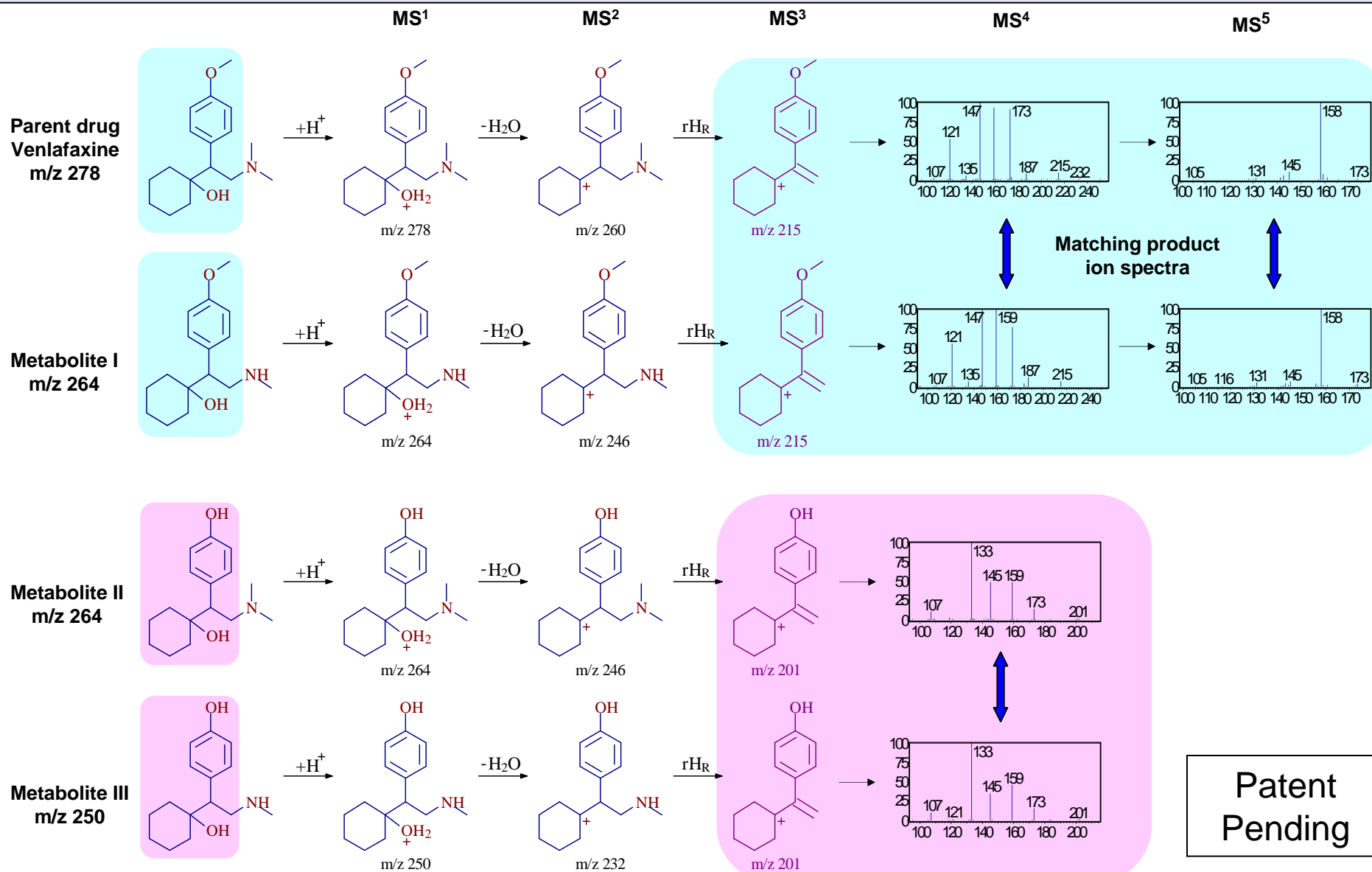


Venlafaxine m/z 278

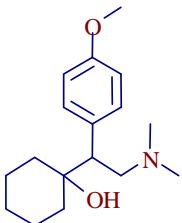
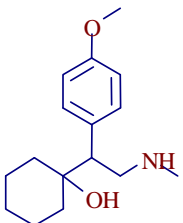
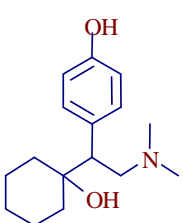
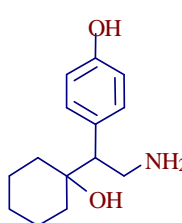
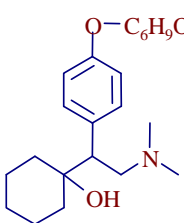
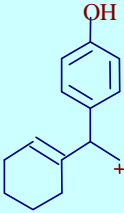
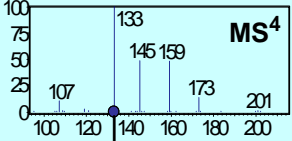
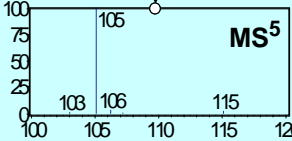
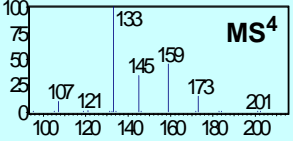

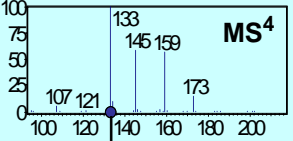
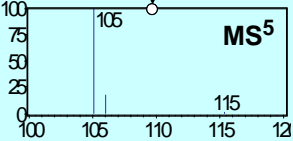
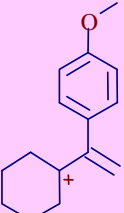
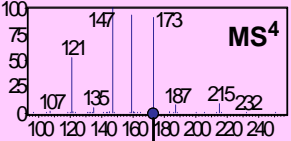
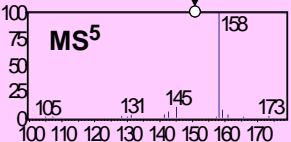
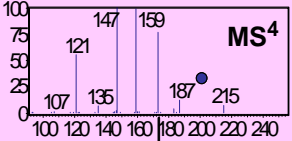
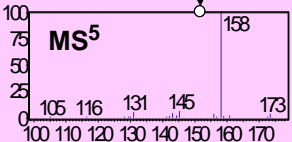
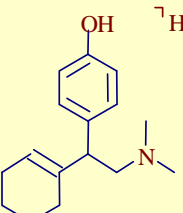
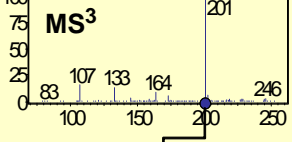
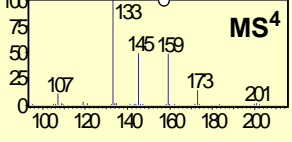
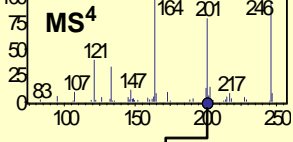
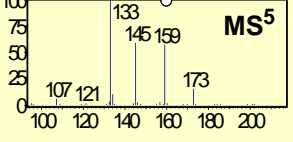
LC-MS of 5  $\mu$ M Venlafaxine + Metabolite Standards

# Metabolite ID using PIF

## Venlafaxine

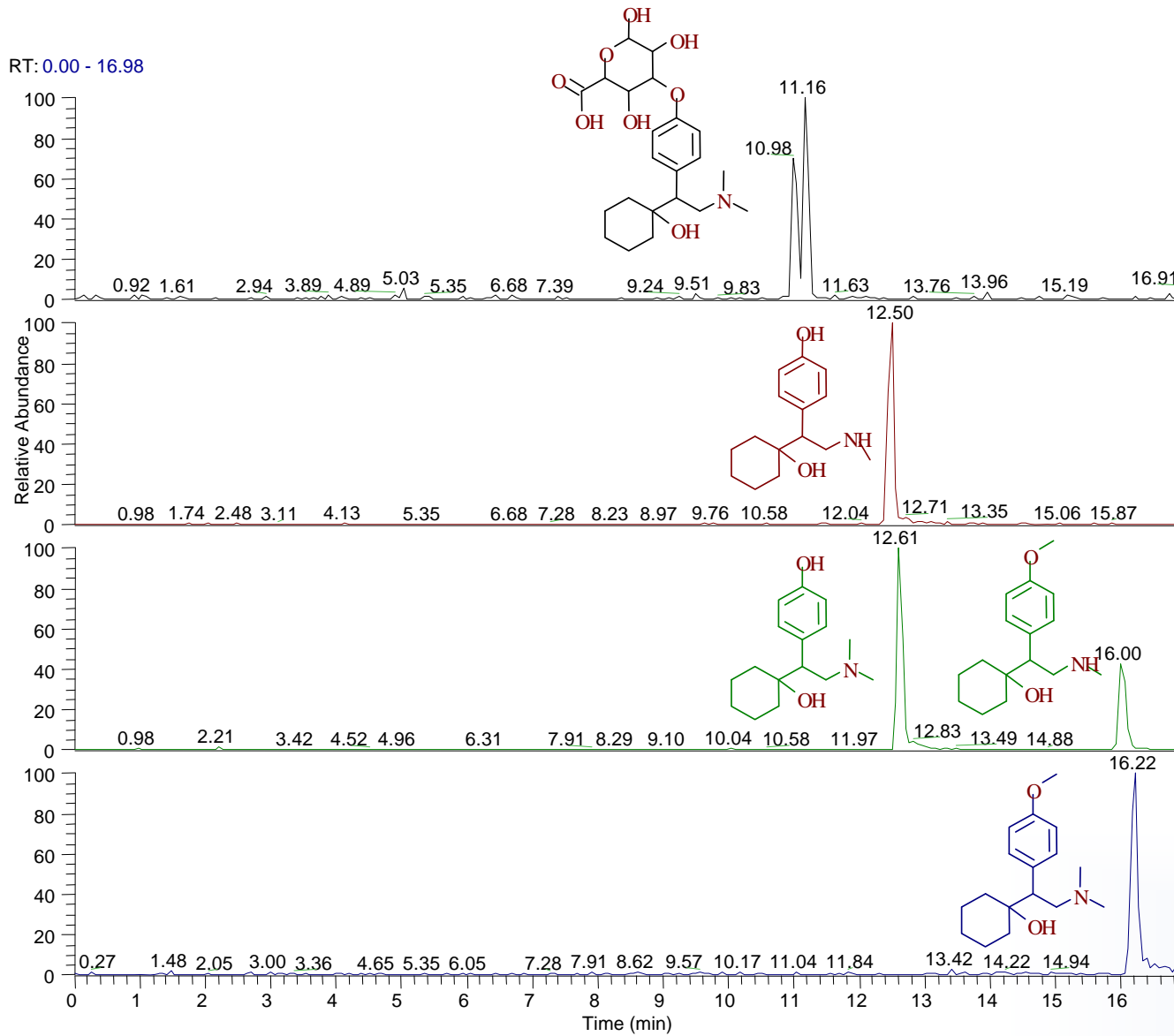


Patent Pending

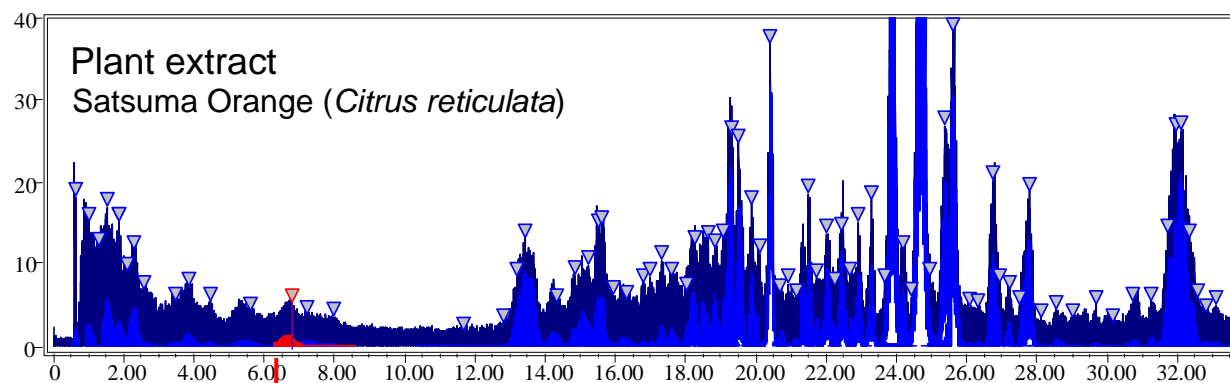
	Venlafaxine	Metabolite I	Metabolite II	Metabolite III	Metabolite IV
<b>Precursor Ion Fingerprinting (PIF)</b>					
 <b>m/z 201</b>			<p><b>precursor ion m/z 201</b></p>  	<p><b>precursor ion m/z 201</b></p>  	<p><b>precursor ion m/z 201</b></p>  
 <b>m/z 215</b>	<p><b>precursor ion m/z 215</b></p>  	<p><b>precursor ion m/z 215</b></p>  			
 <b>m/z 246</b>			<p><b>precursor ion m/z 246</b></p>  		<p><b>precursor ion m/z 246</b></p>  

# Metabolite ID using PIF

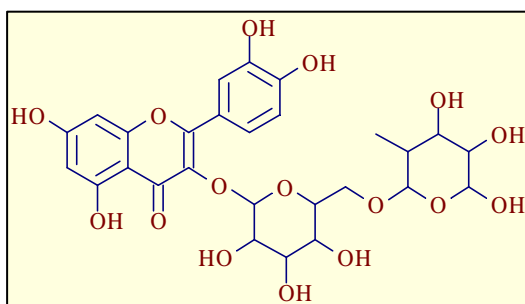
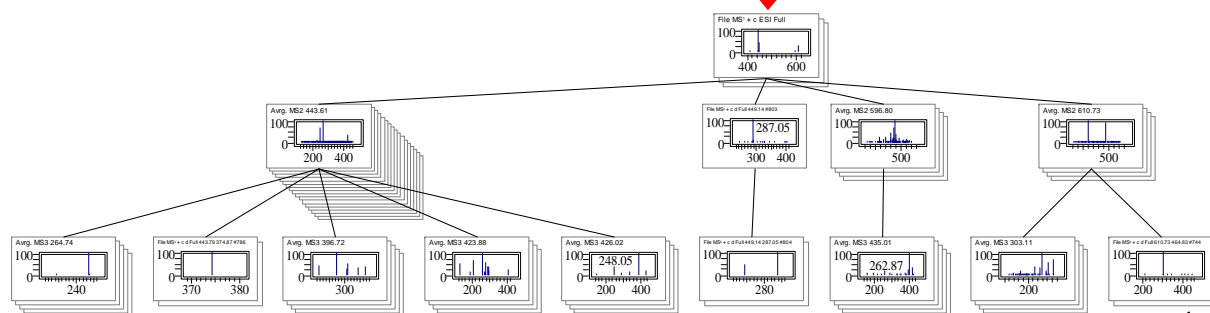
## Venlafaxine



# Flavonoid Metabolite Identification by PIF

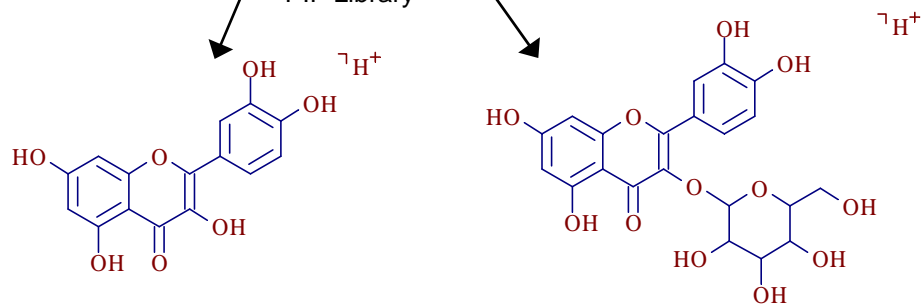


ESI (+), Finnigan LCQ

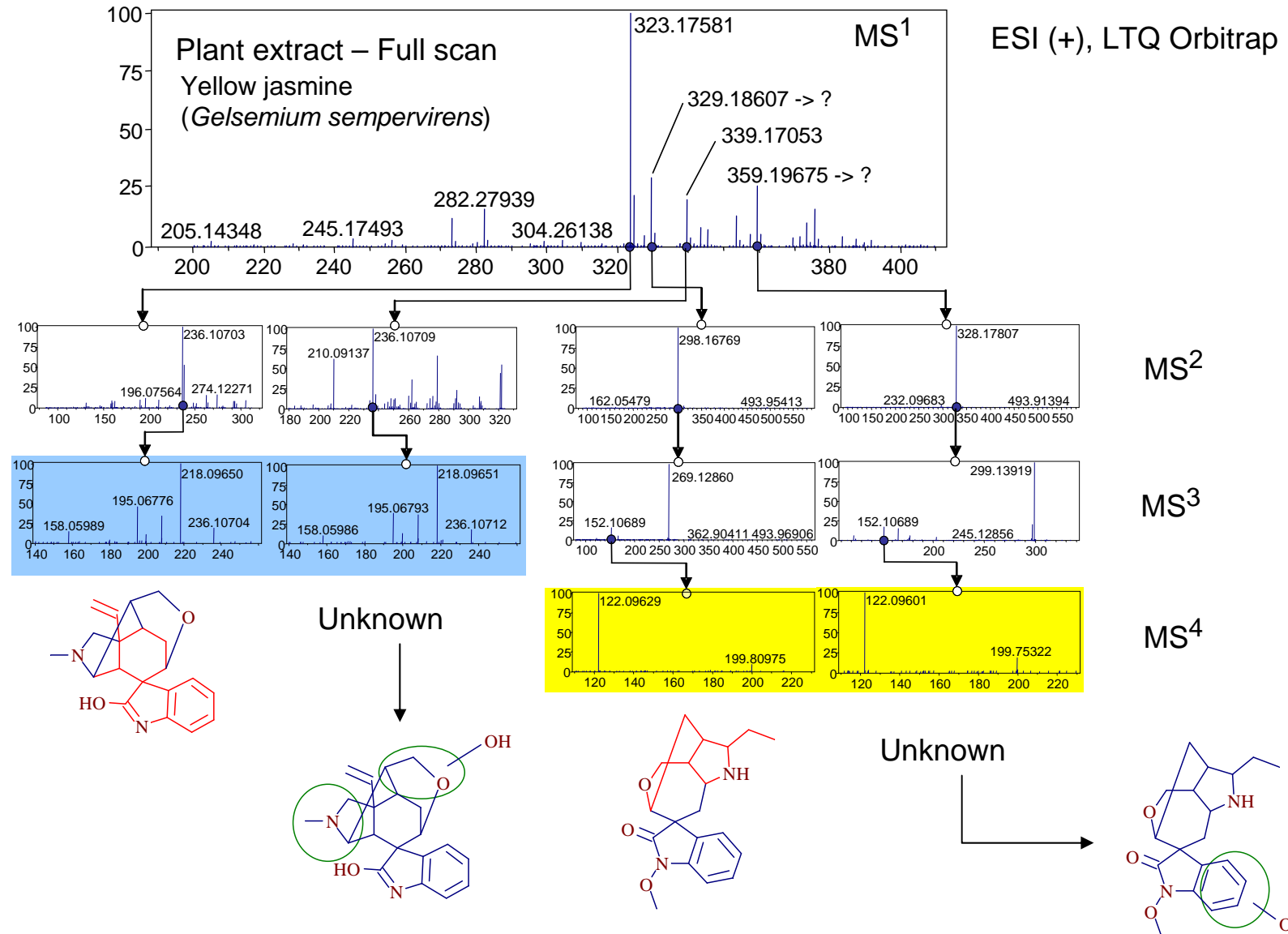


Rutin

PIF Library



# Gelsemine Metabolite Identification by PIF



# PIF Patent



US007197402B2

(12) **United States Patent**  
**Mistrik**

(10) **Patent No.:** **US 7,197,402 B2**

(45) **Date of Patent:** **Mar. 27, 2007**

(54) **DETERMINATION OF MOLECULAR STRUCTURES USING TANDEM MASS SPECTROMETRY**

(75) Inventor: **Robert Mistrik**, Bratislava (SK)

(73) 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.

(21) Appl. No.: **10/967,018**

(22) Filed: **Oct. 14, 2004**

(65) **Prior Publication Data**

US 2006/0085142 A1 Apr. 20, 2006

(51) **Int. Cl.**  
**G01N 31/00** (2006.01)  
**G06F 19/00** (2006.01)

(52) **U.S. Cl.** ..... **702/27; 702/22**

(58) **Field of Classification Search** ..... 702/22,  
702/27, 30, 186; 250/281, 282, 292; 435/6,  
435/7.1

See application file for complete search history.

(56) **References Cited**

#### U.S. PATENT DOCUMENTS

5,072,115 A 12/1991 Zhou ..... 250/281  
5,453,613 A 9/1995 Gray et al. .... 250/281  
6,362,473 B1 3/2002 Germanus ..... 250/282  
6,455,321 B1 9/2002 Chace ..... 436/173  
6,623,935 B2 9/2003 Overney et al. .... 435/7.1  
6,624,408 B1 9/2003 Franzen ..... 250/282  
6,674,069 B1 1/2004 Martin et al. .... 250/287  
6,675,104 B2 1/2004 Paulse et al. .... 702/22  
2002/0063206 A1\* 5/2002 Bateman et al. .... 250/281

2003/0236636 A1 12/2003 Yoshinari et al. .... 702/27  
2004/0229369 A1 11/2004 Kruappa et al. .... 436/86  
2004/0245461 A1\* 12/2004 Senko ..... 250/292

#### FOREIGN PATENT DOCUMENTS

WO WO 02/48699 12/2001

#### OTHER PUBLICATIONS

Gaucher et al., "STAT, A Saccharide Topology Analysis Tool Used in Combination with Tandem Mass Spectrometry", *Anal. Chem.*, vol. 72, pp. 2331-2336, (2000).

\* cited by examiner

*Primary Examiner*—Bryan Bui

(74) *Attorney, Agent, or Firm*—Weingarten, Schurgin, Gagnebin & Lebovici LLP

(57) **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 constructed from 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  
J Am Soc Mass Spectrom. 2009 March;20(3): 370-376  
Michelle T. Sheldon, Robert Mistrik, Timothy R. Croley

Determination of Ion Structures in  
Structurally Related Compounds Using  
Precursor Ion Fingerprinting

Michelle T. Sheldon,<sup>a</sup> Robert Mistrik,<sup>b</sup> and Timothy R. Croley<sup>a</sup>

<sup>a</sup>Commonwealth of Virginia, Department of General Services, Division of Consolidated Laboratory Services, Richmond, Virginia, USA

<sup>b</sup>HighChem, Ltd., Bratislava, Slovakia

Structurally-related alkaloids were analyzed by electrospray ionization/multiple stage mass spectrometry (ESI/MS<sup>n</sup>) at varying collision energies to demonstrate a conceptual algorithm, precursor ion fingerprinting (PIF). PIF is a new approach for interpreting and library-searching ESI mass spectra predicated on the precursor ions of structurally-related compounds and their matching product ion spectra. Multiple-stage mass spectra were compiled and constructed into "spectral trees" that illustrated the compounds' product ion spectra in their respective mass spectral stages. The precursor ions of these alkaloids were characterized and their spectral trees incorporated into an MS<sup>n</sup> library. These data will be used to construct a universal, searchable, and transferable library of MS<sup>n</sup> spectra. In addition, PIF will generate a proposed structural arrangement utilizing previously characterized ion structures, which will assist in the identification of unknown compounds. (J Am Soc Mass Spectrom 2009, 20, 370-376) © 2009 American Society for Mass Spectrometry

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 hard ionization 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 reliably confirm and/or identify compounds [2]. Accordingly, there are a large number of GC/MS instruments available on the market and a comparable number of commercial spectral libraries to go along with them; yet the technique is limited [3] in that many compounds require derivatization to be analyzed by GC/MS [4]. In instances where there are a large number of samples, this can be significantly time-consuming and labor intensive. More importantly, this means that while EI spectral libraries are undoubtedly reliable and effective, they are only applicable to a restricted number of compounds [3].

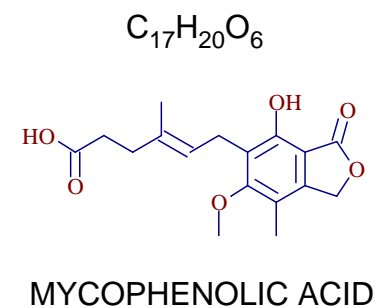
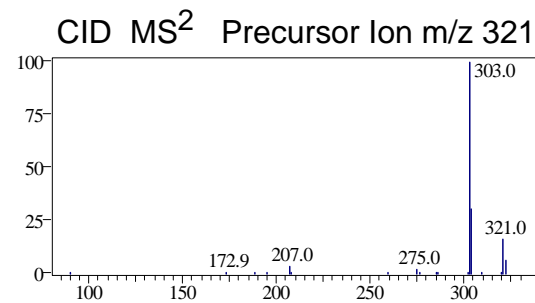
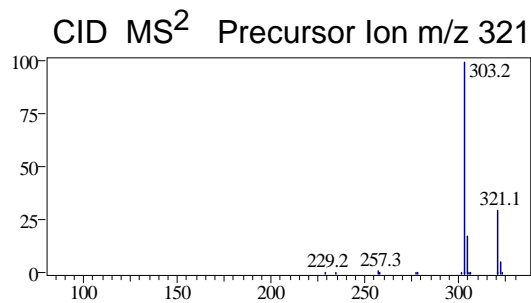
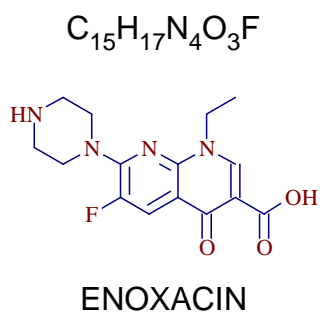
Recently, liquid chromatography/mass spectrometry (LC/MS) employing electrospray ionization (ESI) has gained more use as an effective analytical tool. By

comparison, ESI/MS can be used to analyze more classes of compounds (to include polar compounds), requires less sample preparation (no derivatization step before analysis), is more sensitive, demonstrates increased specificity, and is quantitatively more accurate [3, 5, 6]. In addition, collision induced dissociation (CID) or tandem mass spectrometry (MS/MS) techniques can be incorporated to provoke fragmentation to elucidate considerable structural information pertaining to a compound [5]. Despite all these advantages, to-date there is no universal, searchable fragmentation spectral library that is not confined to a certain type of instrument and/or a specific class of compounds [7].

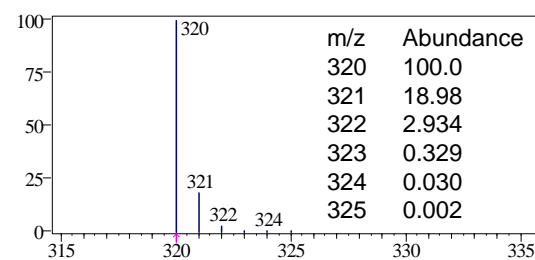
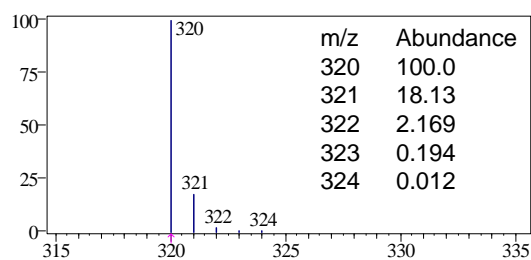
Achieving reproducible ESI/MS<sup>n</sup> data to archive in a spectral library from different instrumental platforms is no easy task. Many have made attempts to utilize already commercially available resources, while even more have tried to develop their own ESI/MS<sup>n</sup> spectral libraries. EI databases, such as the National Institute and Standards and Technology (NIST) and Wiley libraries, are fully evolved and retain an abundant number of compounds. In terms of relations, this would seem like the most reasonable place to start. EI spectra, however in most cases, are very different than those of ESI and fall short of a usable reference database [8]. Prodenhagen et al. 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 export the precursor ion along with the product ion(s) into the NIST database [9]. As a result, only the product ion(s) was/were searchable. Many

Address reprint requests to Dr. T. R. Croley, Commonwealth of Virginia, Department of General Services, Division of Consolidated Laboratory Services, 600 N 5th Street, Richmond VA 23219, USA. E-mail: timcroley@dlgs.virginia.gov

# Spectral Libraries - Fingerprint



## Theoretical Isotopic Pattern

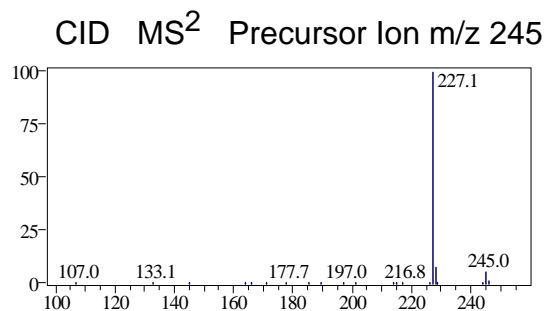
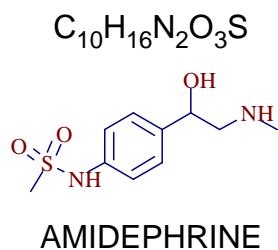


Identical molecular mass (unit), virtually identical isotopic pattern

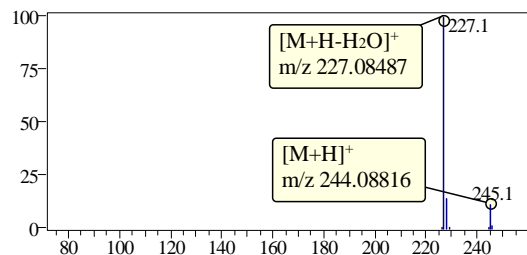
For Confident Compound Identification MS<sup>3</sup> is Needed !!!

# Spectral Libraries - Specificity

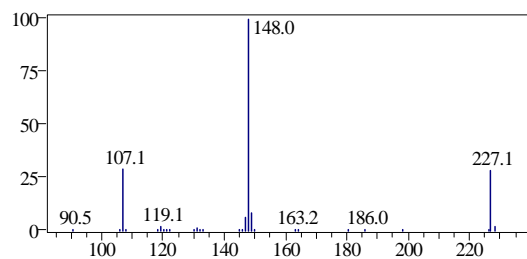
Identical Molecular Formula = Identical Exact Mass, Identical MS<sup>2</sup>



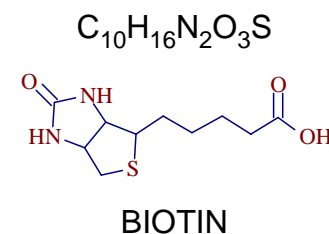
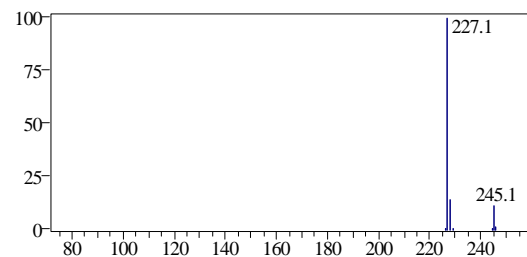
MS<sup>2</sup> Wide band activation



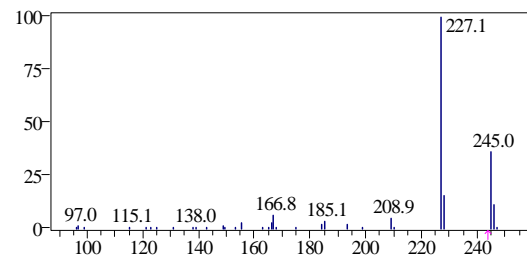
CID MS<sup>3</sup> Precursor Ion m/z 227



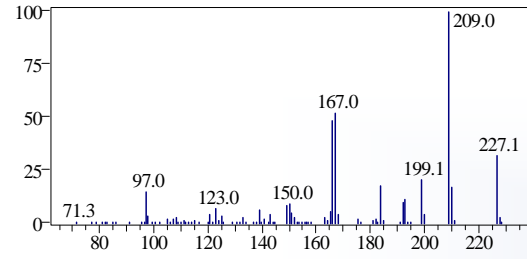
CID MS<sup>2</sup> Precursor Ion m/z 245



MS<sup>2</sup> Wide band activation

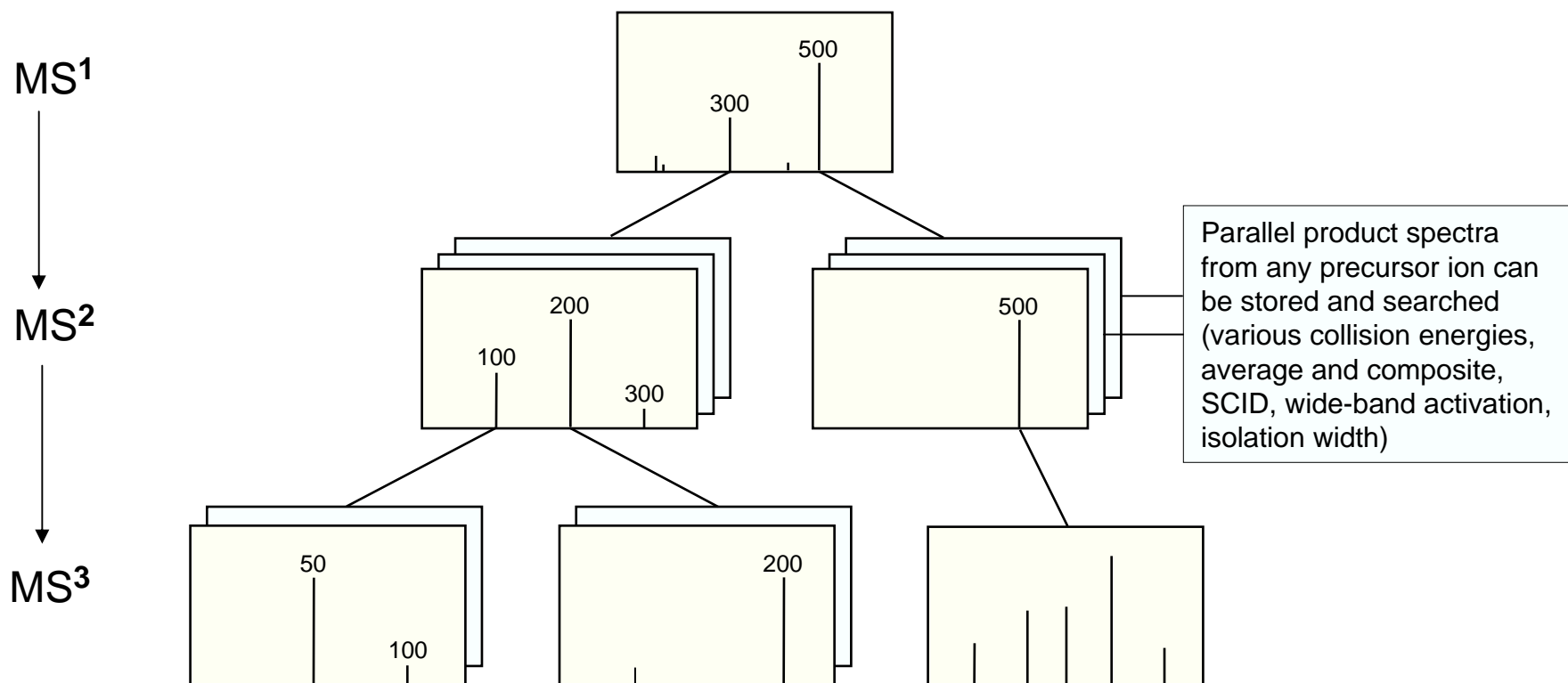


CID MS<sup>3</sup> Precursor Ion m/z 227



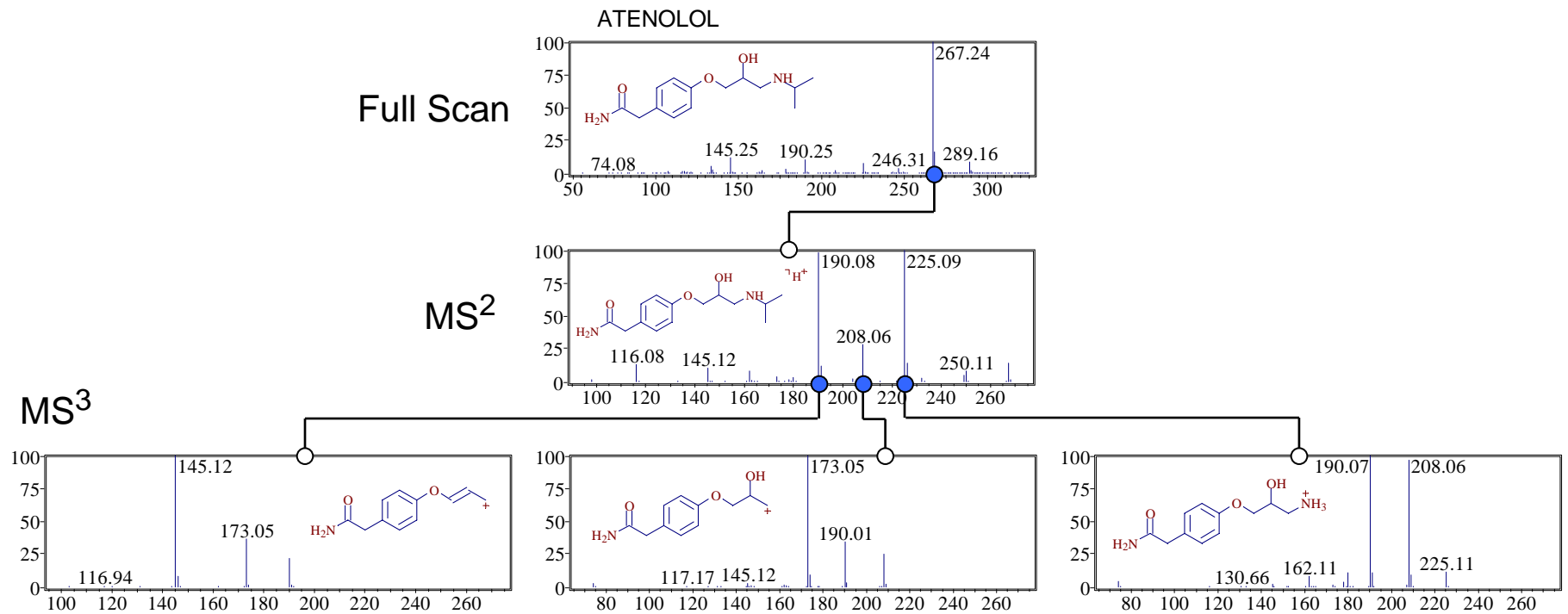
## New MS<sup>n</sup> spectra representation

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