MS Reference Libraries for Forensics: Past, Present and Future

Forensics@NIST 2012
Steve Stein et al.
NIST MS Data Center
Identification
A Central Task in Forensics

• People
  – DNA, Fingerprints, Features, ...  

• Objects
  – Clothing, Weapon, ...  

• Chemicals
  – Molecular Identity
Outline

• Library Background
• Nature of the Data
• Identification by MS
• NIST Tools
• Tandem MS
• Future
Library Background
Evolution of the NIST MS Library

- 1970’s: NIH/EPA Collection of Collections
  - Fales, Heller

- 1980’s: To EPA Cincinnati
  - Budde


- 2000’s: Tandem MS GC Retention

- 2010’s: To NIST PC-XT Version

- Red Books 9-track Tape 300 Baud Modem

- Evaluated Library AMDIS

- Peptides High Resolution
Numbers of EI Spectra

- Replicates
- Compounds
Data Sources

• In the Beginning: Library of Libraries + Literature
• Contractor Labs
• NIST Measurements
• Contributors
  – Industry, Academics, Organizations, Crime Labs, ...

• New Spectra (ca. 10,000 / year)
  – Derivatives of common chemicals
  – Metabolites (human and plant)
  – Environmental/Security
  – Newly regulated compounds
SWGDRUG Mass Spectral Library

SWGDRUG has compiled a mass spectral library from a variety of sources, containing drugs and drug-related compounds. All spectra were collected using electron impact mass spectrometry systems. This library is available for download from this website.

DISCLAIMER: Although SWGDRUG makes an effort to review the accuracy of spectra prior to entry, this library should only be used as an analytical tool. SWGDRUG recommends the use of traceable reference materials to support identifications of drugs (Part IV B – Quality Assurance Section 2.3)

The SWGDRUG library is supported by the NIST MSSEARCH program, which is available on-line at no charge (see below). Additionally, the library was converted to Agilent Technologies format. Lastly, two raw data formats are included below depending upon your desired application. Click on the appropriate link below to download the compressed file and follow the instructions below.

<table>
<thead>
<tr>
<th>SWGDRUG MS Library</th>
<th>Version 1.5 (July 11, 2012)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIST Format</td>
<td>Agilent Format</td>
</tr>
<tr>
<td>JCAMP Format (.jnp)</td>
<td>Compound List</td>
</tr>
<tr>
<td>Shimadzu Format</td>
<td></td>
</tr>
</tbody>
</table>
The Critical Evaluation of a Comprehensive Mass Spectral Library

P. Ausloos, C. L. Clifton, S. G. Lias, A. I. Mikaya, S. E. Stein, and D. V. Tchekhovskoi
NIST Mass Spectrometry Data Center, Gaithersburg, Maryland, USA

O. D. Sparkman
Sparkman and Associates, Antioch, California, USA

V. Zaikin
Topchiev Institute of Petrochemical Synthesis, Moscow, Russia

Damo Zhu
Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, China

• Initial Manual Evaluation
• Spectrum + Structure Computer Processing
• Chief Evaluator: Mark as Best, Alternate, Reject
• Add to Archive → Build Library
Nature of the Data
"The Decomposition of Hydrocarbons in the Positive Ray Tube"

H.R. Stewart & A.R. Olson, 1931

Fig. 3.—Typical runs with propane and butane. Ordinates represent electrometer deflections; abscissas, current through the magnet. Numbers given are calculated masses of corresponding singly charged ions.
A mass spectrum is a property of an ion
Identification by MS
Identification by Pattern Matching

• Mass spectra are molecular ‘properties’
  – Reflect molecular structure

• Peaks are easily formed stable fragments
  – May not be unique to compound
Identification by GC/MS

Match EI Mass Spectra and Retention Time

Compute Score

But, Identification is Indirect and Depends on the Analyte
Bayes Rule

Prior Probability:
Before Experiment

False Negative Potential

False Positive Potential

Final Confidence

Starting Confidence

Change in Confidence

* Odds Version


P ( ID | Score )

P ( FP | Score )

Analyte is Identified Correctly

Mass Spectral Reference Libraries for Chemical Identification

Influence of Library Search

P ( ID | Score )

P ( FP | Score )

P ( Score | ID)

P ( Score | FP)
Class Identification or False Positive?
Mass Spectrometry Tools

This site provides access to chemical data products and updates from NIST. Currently AMDIS (Automated Mass Spectral Deconvolution and Identification System), Mass Spectrum Interpreter, and the Mass Spectral Digitizer Program are available as well as the Peptide Databases and a demo version of the NIST08 library and the NIST MS Search Program v.2 of from this Web site.

Software

- NIST/EPA/NIH Mass Spectral Database – the product of a multiyear comprehensive evaluation and expansion of the world’s most widely used mass spectral reference library
- NIST MS Search Program – what’s new, software updates, demo, documentation and support programs
- NIST 11 GC Method / Retention Index Database – information about this new product
- NIST 12 MS/MS Database – information about this new product **NEW**
- Mass Spectrum Interpreter – information and downloads for version 2 of this program which predicts fragmentation based on structures and correlates with mass spectra (EI and MS/MS)
- AMDIS – computer program that extracts spectra for individual components in a GC/MS data file
- Mass Spectrum Digitizer Program – a tutorial on how to use the program (includes program download) that allows the digitization of graphical spectra

Publicly Available Libraries

- NIST Peptide Databases – These databases contain nearly 1 million MS/MS Spectra for multiple biological species
- EPA Tandem Mass Spectrometry Library – Starter Library of Environmentally Relevant Compounds
- DART Forensics Library – A library of in-source fragmentation spectra of DART-generated ions

Other Products and Services

- Where to Get NIST Databases – Databases are available only through distributors. See individual product descriptions for lists of distributors.
- Presentations and Publications – List of publications and presentations
- Chemical Reference Data Group – Home page of the group responsible for this site
- NIST Chemistry WebBook – Site with a diverse collection of data including electron ionization (EI) mass spectra

chemdata.nist.gov
Traditional Library Search

Search List

Hit List

Score Histogram

Query Spectrum

Library Spectrum

2011 Version - 213K EI, 5K CID, 71K RI Compounds
Substructure Analysis

Chemical Substructure Identification by Mass-Spectral Library Searching
JASMS 6 (8) 644-655 (1995)
MS Interpreter
Automated Mass Spectral Deconvolution and Identification System
Created for Chemical Weapons Treaty Verification: “Blinded CW Identification”

JASMS 1999 10 770-781
peptide.nist.gov

NISTMSQC: Full Analysis of LC-MS/MS data

Library/quality metrics

Tandem MS
**NIST Tandem Mass Spectral Library 2012**

### Classes:
Metabolites, Drugs, Sugars, Phospholipids, Peptides, Surfactants, etc.

### Precursors:
- $[M+H]^+$
- $[M+2H]^{2+}$
- $[M-H]^{-}$
- $[M+Na]^+$
- $[M+NH4]^+$
- $[Cat]^+$
- $[An]^{-}$
- $[p-H2O]$
- $[p-NH3]$, etc.

### New Software Features:
- Exact or isotopic precursor mass & fragment ions.
- Formats: *mzXML, mzData, mgf, msp, dta, pkl, JCAMP*, etc.
- Compatible with NIST EI & Peptide Tandem Libraries.
- New methods for finding targets in the presence of noise.

### New Scoring:
- Compounds with few dominant peaks.
- Compensates for $m/z$ tuning errors.

### Fragmentation Type vs. Precursor Ions:
- Ion Trap: >10,000
- Beam Collision Cell (QTOF, QQQ, HCD): >8,000

### Statistics:
- Compounds: 7,020
- Precursor Ions: 15,517
- Spectra: 123,781
Emerging MS Methods
3,362 DART CID Spectra, 757 Compounds

Robert L. Steiner, Virginia Crime Lab
Chip Cody, JEOL
http://chemdata.nist.gov/
Future
Future Work

• Algorithms
  – Accurate ID confidence

• ‘Recurrent’ Spectrum Libraries
  – Combine with IDs for all mixture components
  – Substance-based libraries

• SRM/D
  – Reference Materials + Reference Data
Algorithms
with Wallace, Kearsley, Allison @NIST

• ‘Dot Product’ Function is Best Measure of Spectrum Similarity

• Using Spectrum Similarity Only Ignores:
  – Chemical/Spectrum Class
  – ‘Prior Probability’

• Secondary Scoring is Promising
  – Use spectrum/compound class to re-score
  – Adjust for Sample/Method

• Target: Identification Probability with Error Limits
Goal: Interpret All Spectra

- GC/MS: Begin with AMDIS
  - Chemdata.NIST.Gov
- LC-MS/MS: Begin with NISTMS QC
  - Peptide.NIST.Gov

Goal: Interpret All Spectra
- Classify Each Spectrum:
  - Identified
  - ‘Recurrent’ spectrum
  - Unknown compound
  - Mixture
  - Noise/Background
• **Standard Reference Data + Data**
  - MS library, Thermodynamic Data, Chemistry Webbook, ...

• **Standard Reference Materials + Data**
  - Substance + Data + Libraries

SRM/D.NIST.GOV