

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

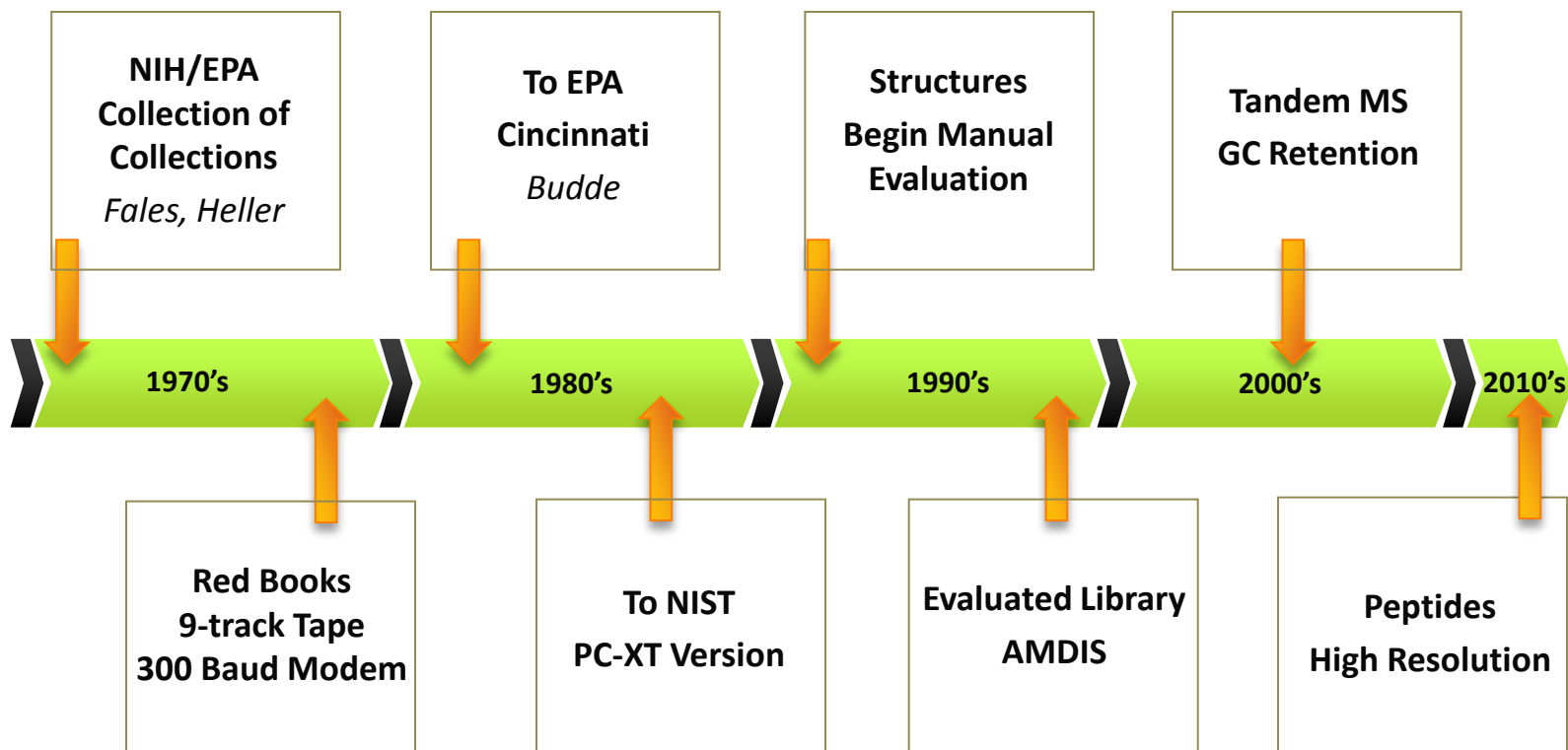
- People
 - DNA, Fingerprints, Features, ...
- Objects
 - Clothing, Weapon, ...
- Chemicals
 - Molecular Identity

Outline

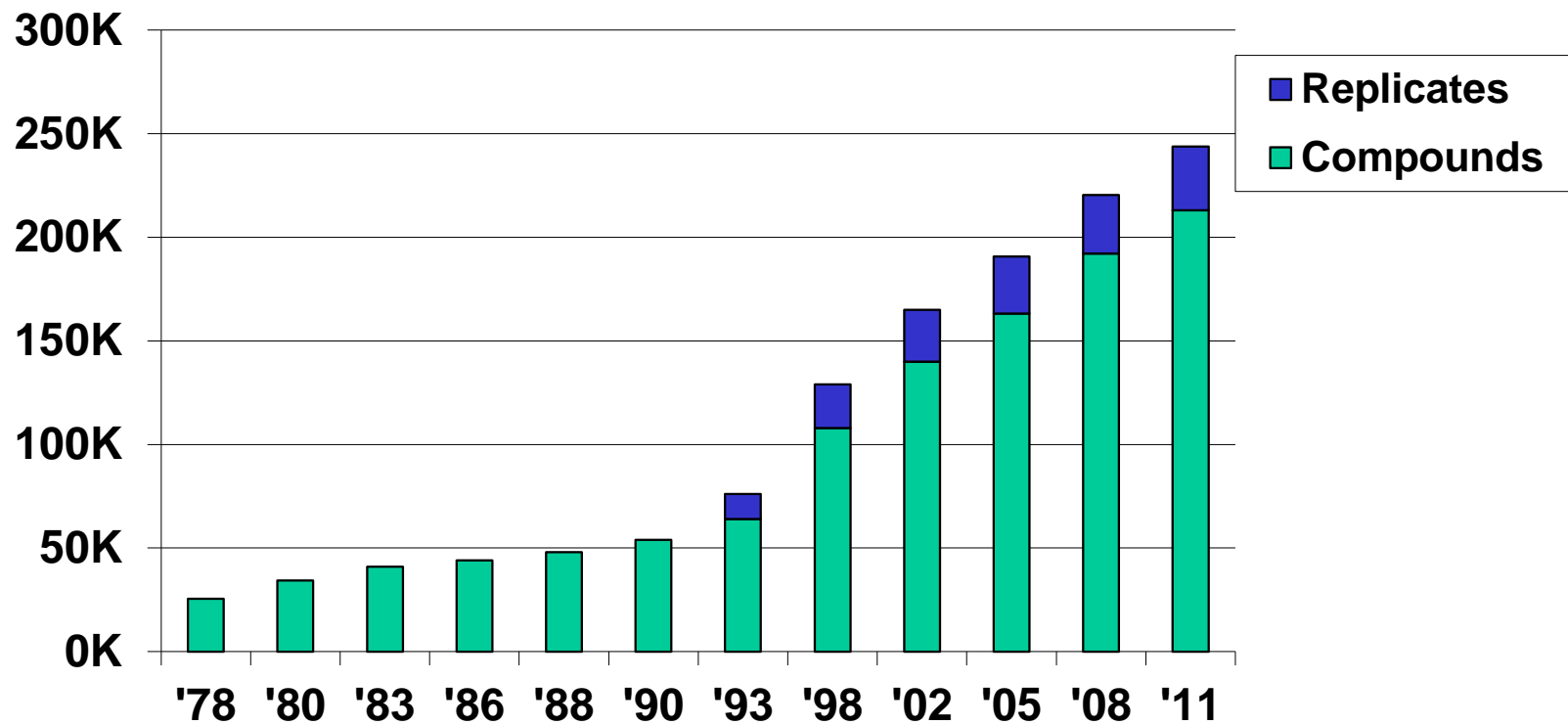
- Library Background
- Nature of the Data
- Identification by MS
- NIST Tools
- Tandem MS
- Future

Library Background

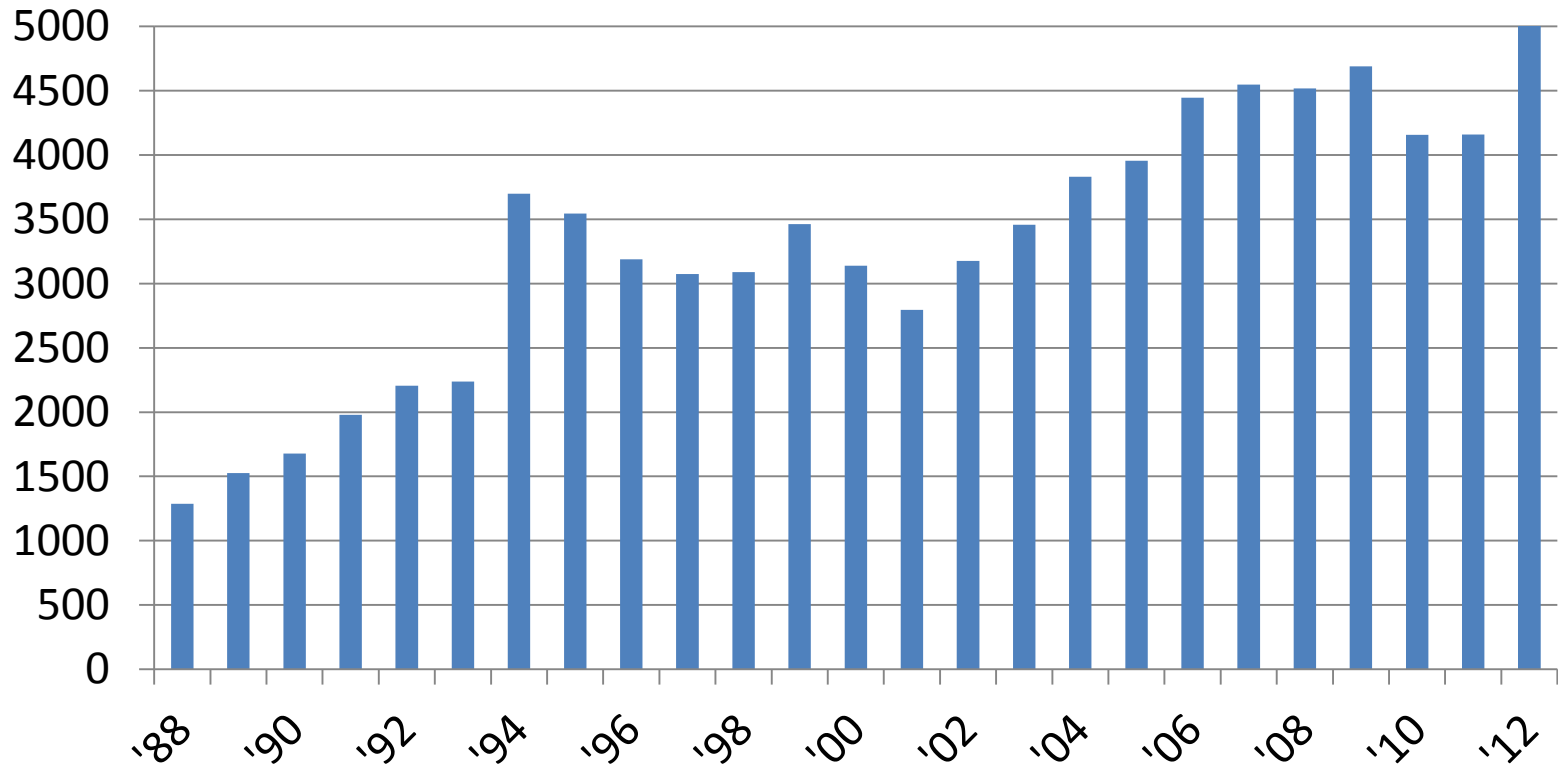
Evolution of the NIST MS Library



Numbers of EI Spectra



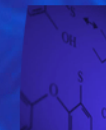
EI Libraries Distributed/Year



50 Distributors

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



Home

About Us

Documents

Membership

Meetings

Resources



WHAT'S NEW





- 1) Edition 6.0 of [SWGDRUG Recommendations](#) which includes a new chapter Part IIIC for the Analysis of Clandestine Drug Laboratory Evidence
- 2) Searchable Mass Spectral Library [Version 1.5](#) dated July 11, 2012
- 3) New Draft [Supplemental Document \(SD-4\)](#) - Provide your comments by September 12, 2012
- 4) Addition of [Sampling Probability Calculator](#)

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.

SWGDRUG MS Library Version 1.5 (July 11, 2012)	
NIST Format	Agilent Format
	
JCAMP Format (.hpi)	Compound List
	
Shimadzu Format	

“Evaluation”

ACCOUNT AND PERSPECTIVE

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

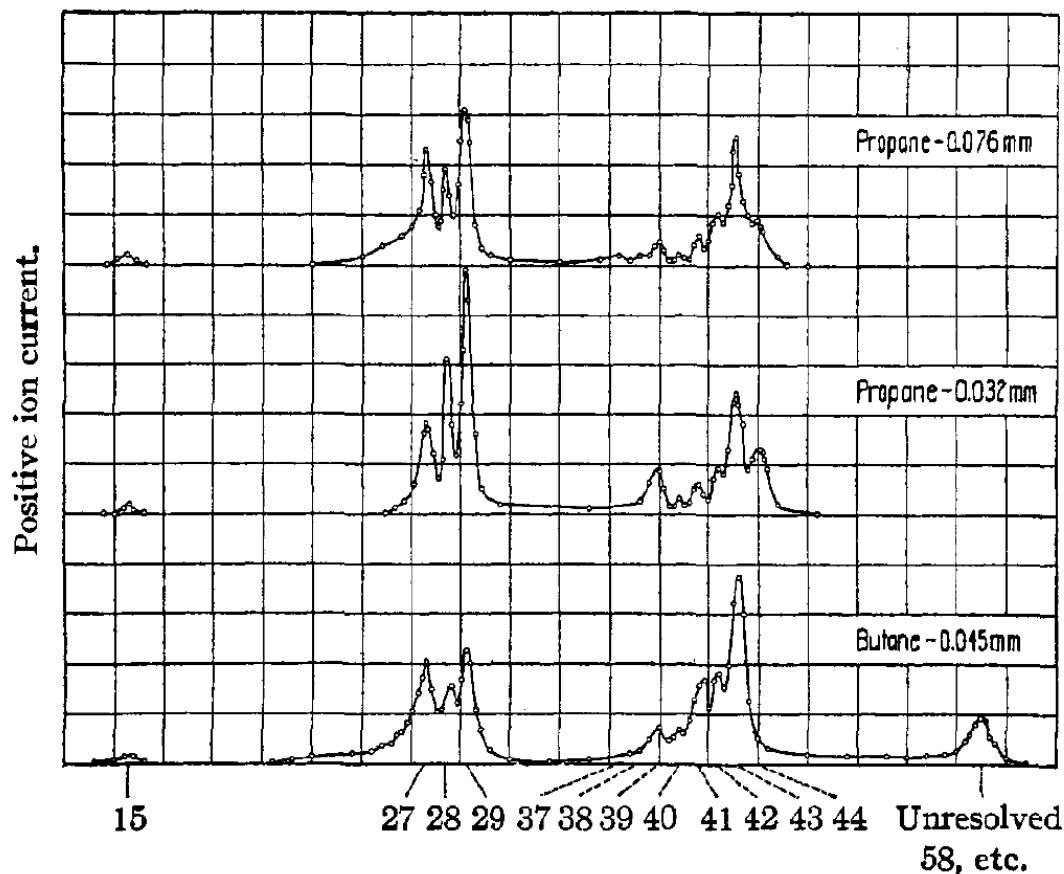
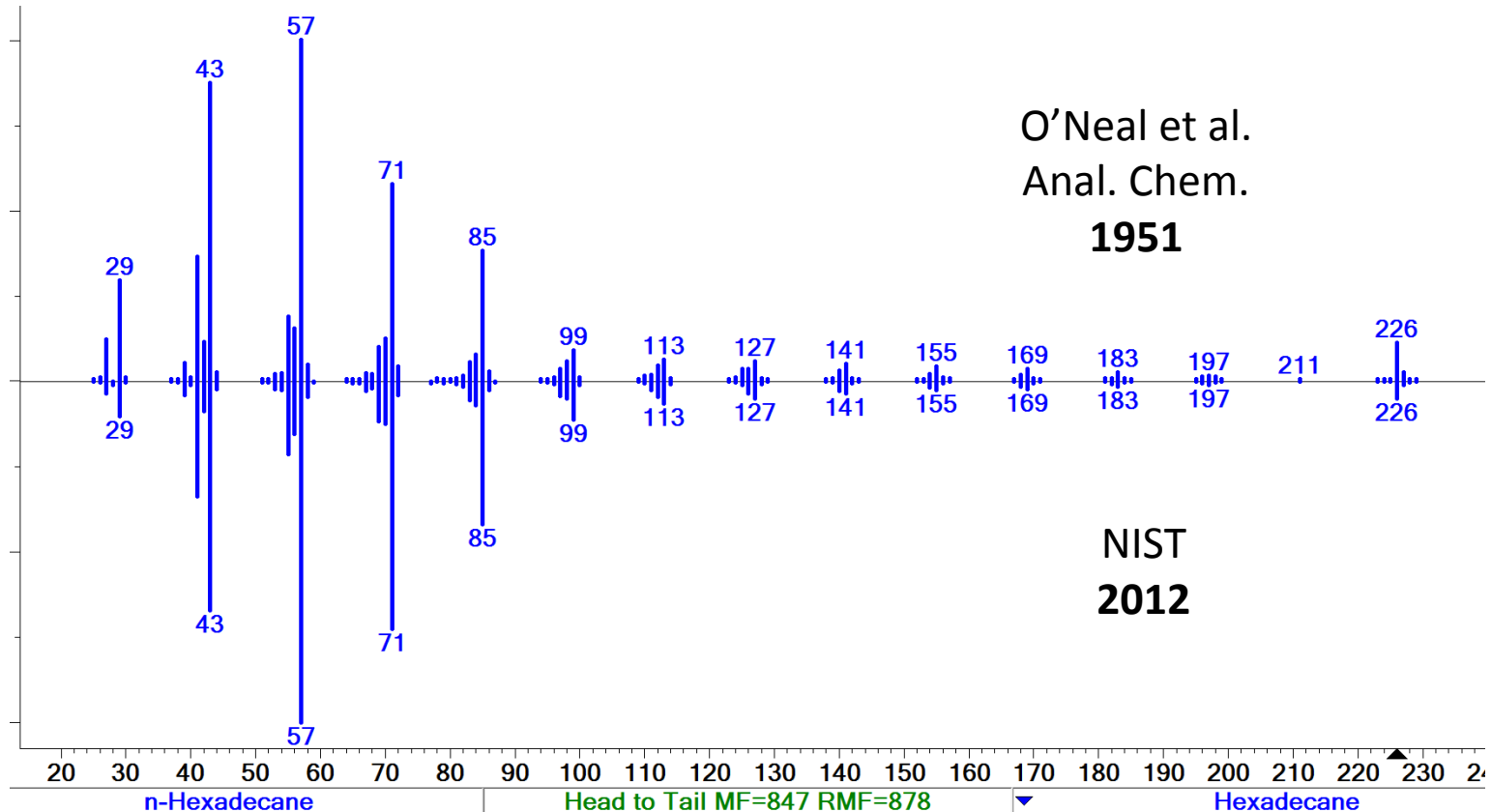


Fig. 3.—Typical runs with propane and butane. Ordinates represent electrometer deflections; abscissas, current through the magnet. Numbers given are calculated masses of corresponding singly charged ions.

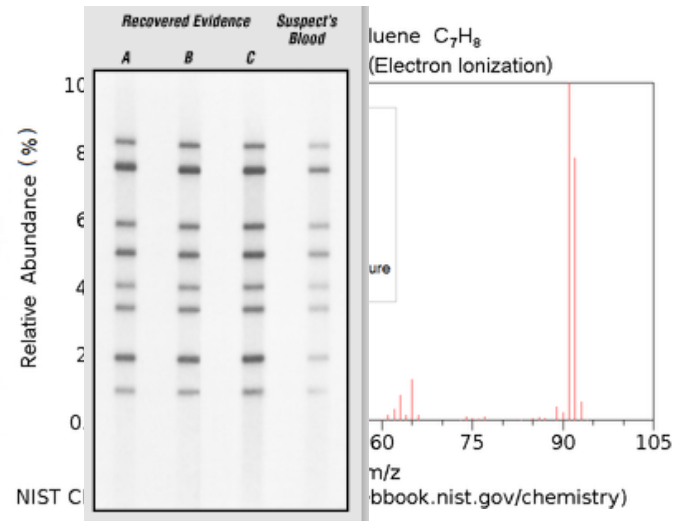
Mass Spectra are Reproducible



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

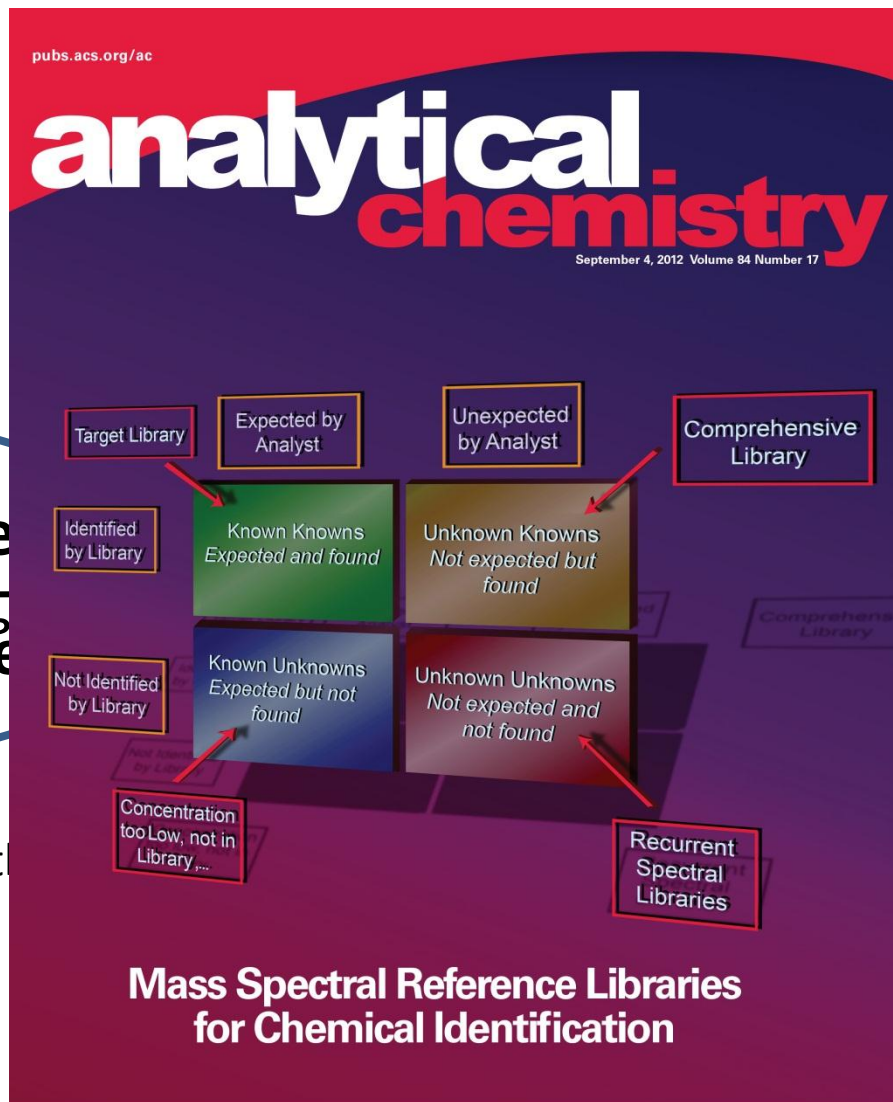
Identification by GC/MS

Match EI Mass Spectra
and Retention Time

Compute Score



But, Identification is Indirect and Depends
on the Analyte



$$\frac{P(\text{ID} | \text{Score})}{P(\text{FP} | \text{Score})}$$

Find
Confidence

Analyte is Identified Correct

False Negative Potential

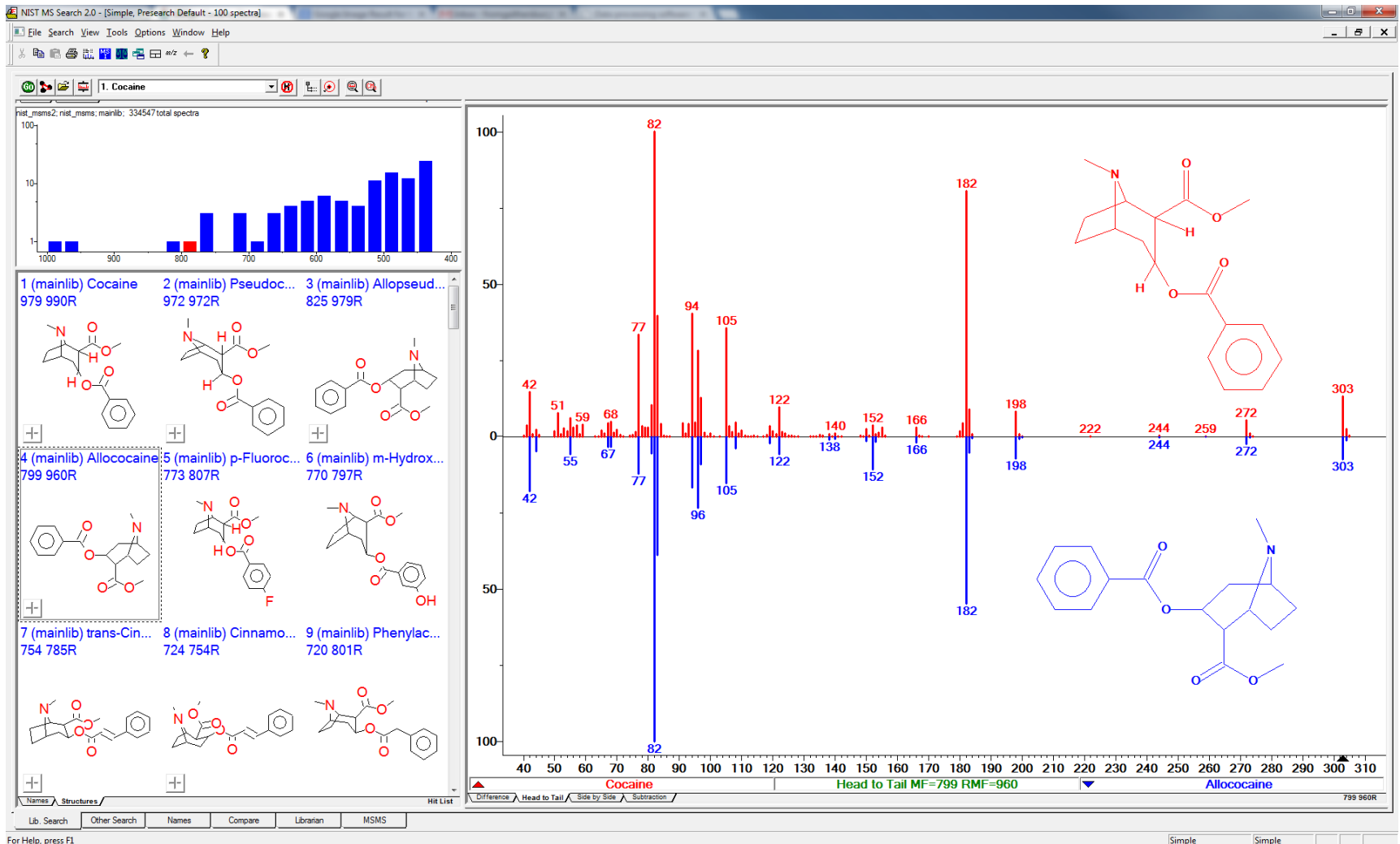
$$\frac{P(\text{Score} | \text{ID})}{P(\text{Score} | \text{FP})}$$

change in
Confidence

False Positive Influence of Library Search

Olds Version


Class Identification or False Positive?



Boehner rej | x 1.What is In | x 2.How did | x 3.What use | x NIST www.nist.g | x Kill the doll | x NIST Chemical R | x NIST Mass Spectr | x

chemdata.nist.gov

NIST
National Institute of
Standards and Technology



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 NIST 08 Library and the NIST MS Search Program v.2.0f 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

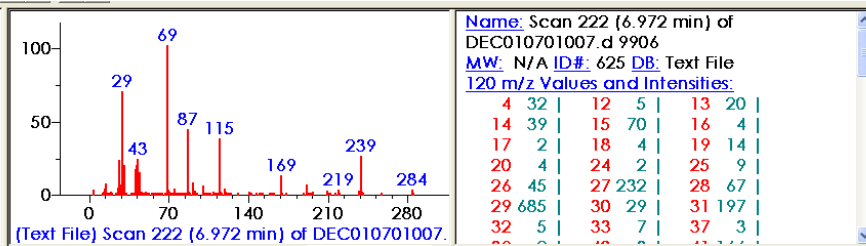
chemdata.nist.gov

Traditional Library Search

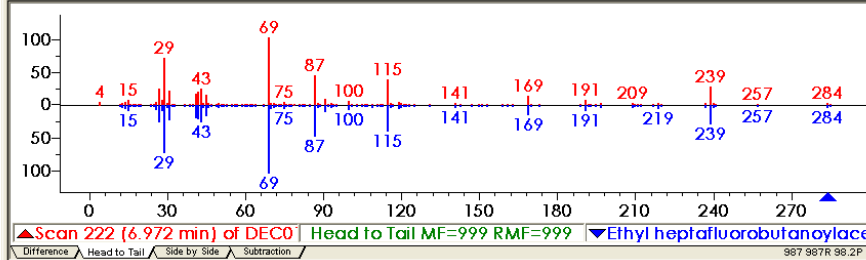
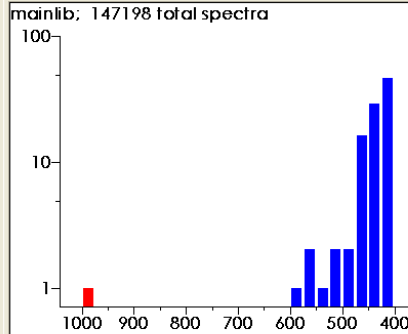
Search List

#	S...	Name
4	A	Scan 1339 (26.715 min) of JAN21
5	A	Scan 1889 (36.437 min) of JAN25
6	A	Scan 1108 (22.633 min) of JAN24
7	A	Scan 222 (6.972 min) of DEC0107
8	A	Scan 815 (17.453 min) of DEC07
9	A	Scan 741 (16.144 min) of DEC10
10	A	Average of 38.980 to 39.086 min
11	A	Scan 1038 (21.397 min) of MAR0

Query Spectrum



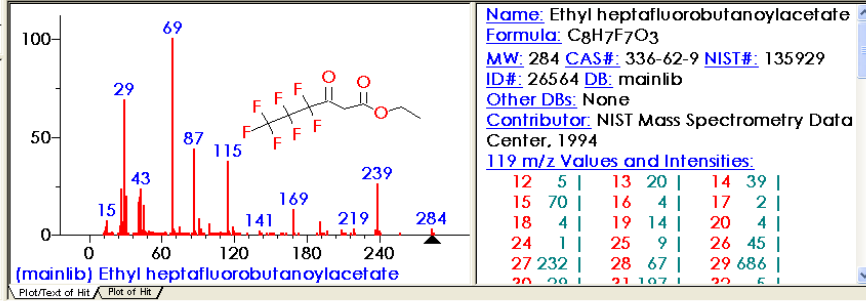
Score Histogram



Hit List

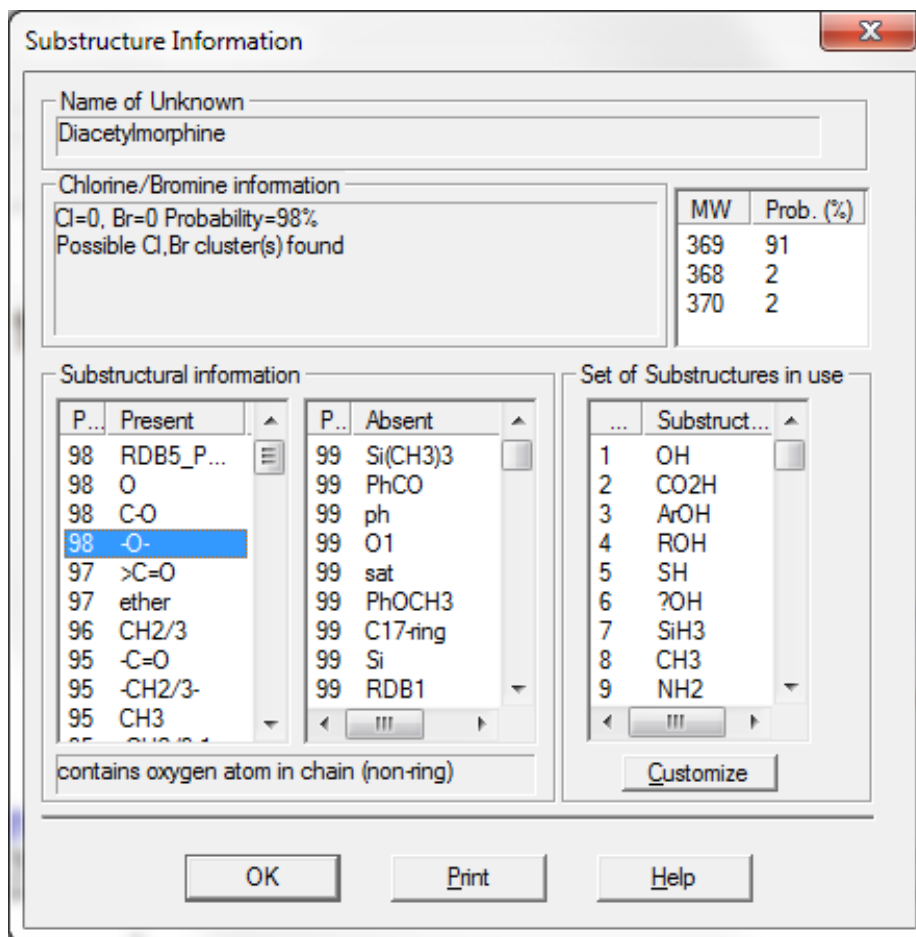
...	Li...	Ma...	R.Maf...	Prob...	Name
1	M	987	987	98.2	Ethyl hept
2	M	600	720	0.98	Ethyl pent
3	M	567	576	0.25	4,6-Nonan
4	M	554	615	0.16	1,1,1,5,5,6
5	M	543	641	0.11	Butanoic c
6	M	504	690	0.02	1,3-Dioxar
7	M	502	683	0.02	1,3-Dioxar
8	M	499	683	0.02	1,3-Dioxar

Library Spectrum



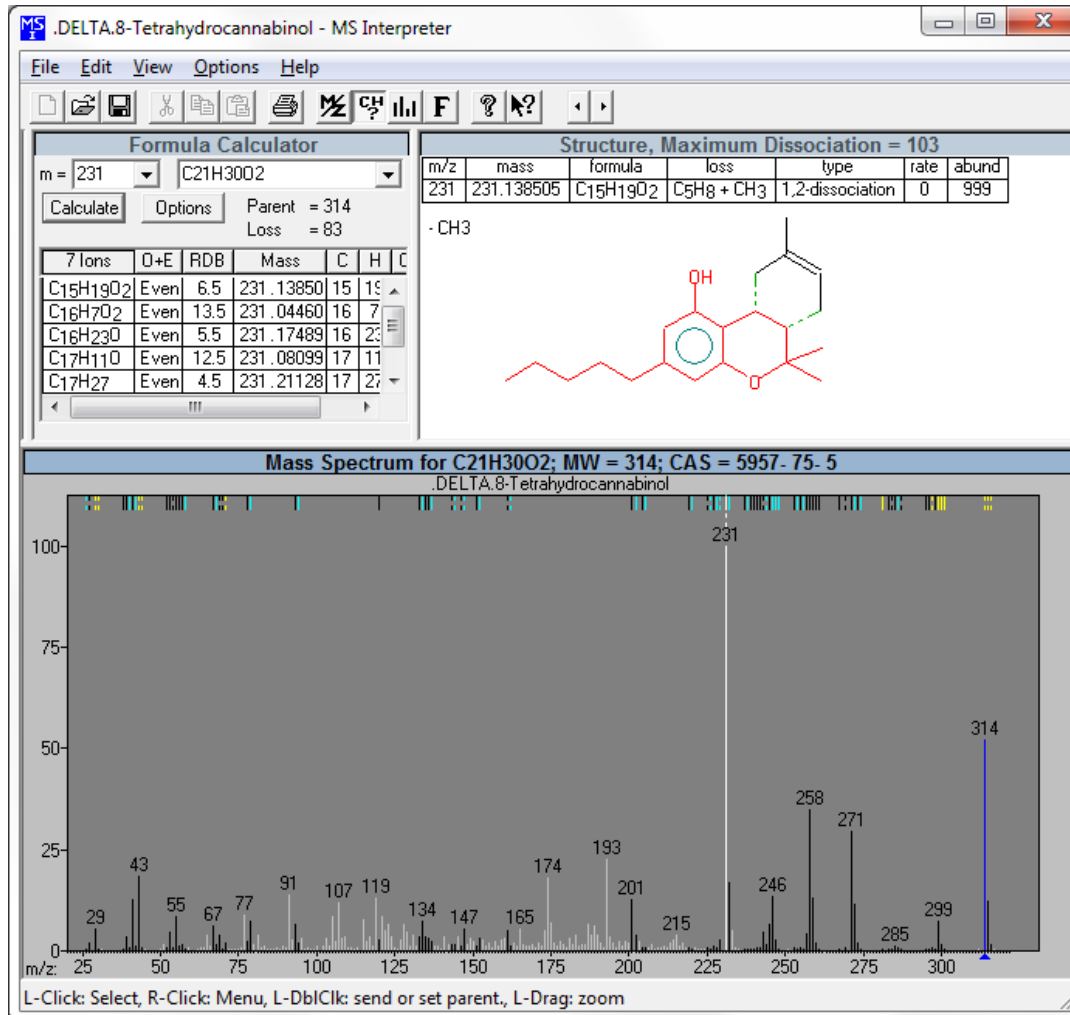
2011 Version - 213K EI, 5K CID, 71K RI Compounds

Substructure Analysis

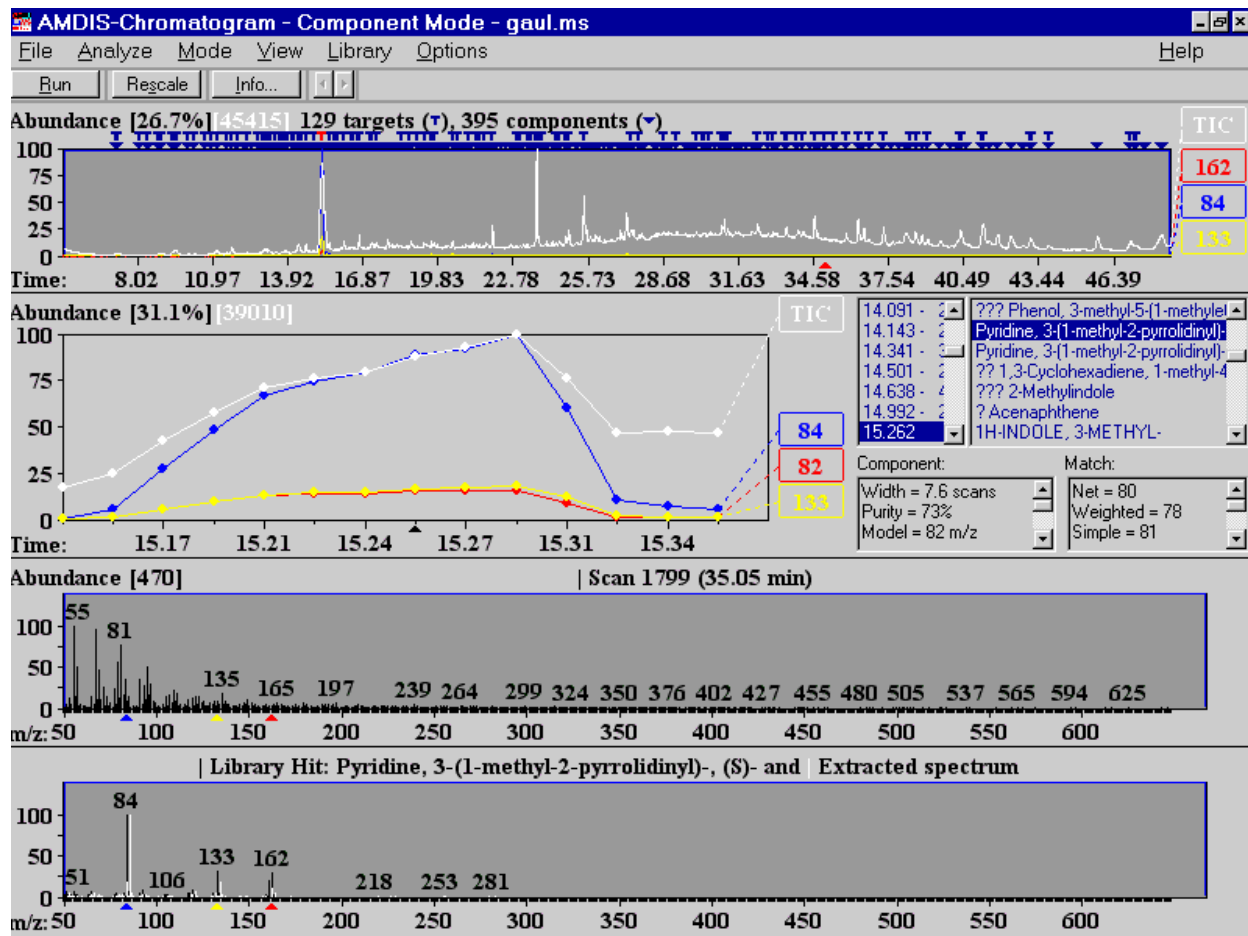


Chemical Substructure Identification by Mass-Spectral Library Searching
JASMS 6 (8) 644-655 (1995)

MS Interpreter

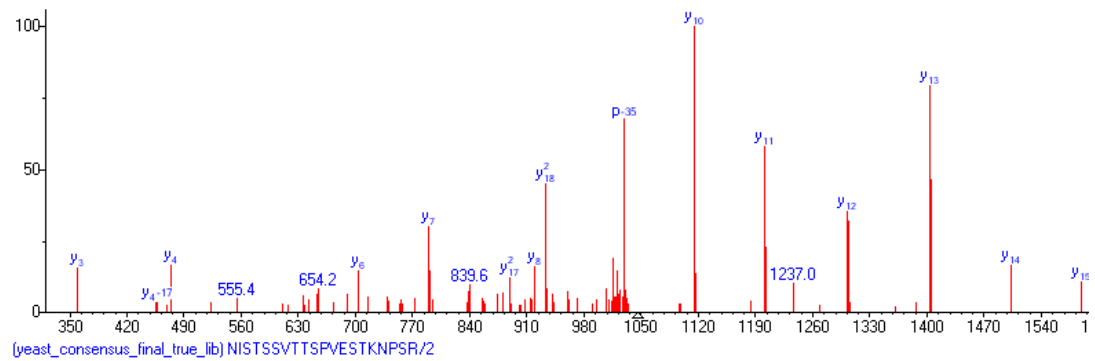


AMDIS



Automated Mass Spectral Deconvolution and Identification System
Created for Chemical Weapons Treaty Verification: “Blinded CW Identification”
JASMS 1999 10 770-781

peptide.nist.gov



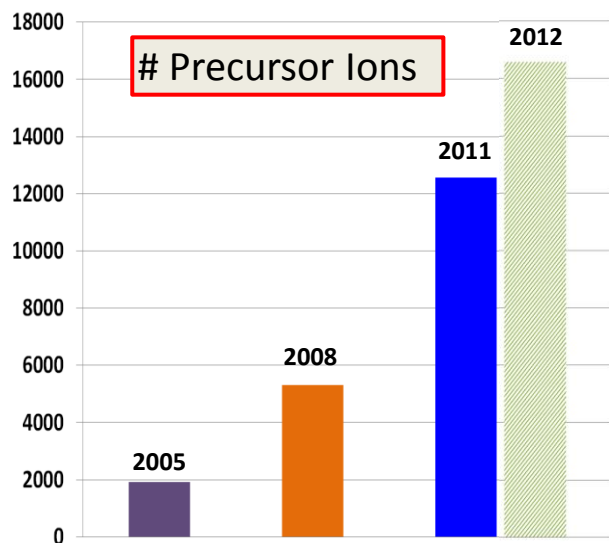
Build Date	Download*	Seq. File†	Library	Organism	Instrument	Exps.‡	Spectra	Peptides	Coverage
By species									
Feb. 04, 2009	ASCII NIST SpectraST FASTA		human	<i>H. sapiens</i>	it	348	261,778	158,522	16%
Jul. 22, 2008	ASCII NIST SpectraST FASTA		human	<i>H. sapiens</i>	qtof	61	12,473	10,139	1%
Apr. 29, 2009	ASCII NIST SpectraST FASTA		mouse	<i>M. musculus</i>	it	117	131,628	78,613	8%
Jul. 14, 2008	ASCII NIST SpectraST FASTA		drosophila	<i>D. melanogaster</i>	it	97	96,542	62,162	13%
Jun. 26, 2009	ASCII NIST SpectraST FASTA		C.elegans	<i>C.elegans</i>	it	2	81,177	49,754	12%
May 04, 2009	ASCII NIST SpectraST FASTA		yeast	<i>S. cerevisiae</i>	it	63	87,676	52,076	23%
May 06, 2009	ASCII NIST SpectraST FASTA		yeast	<i>S. cerevisiae</i>	qtof	5	3,176	2,960	2%
May 21, 2009	ASCII NIST SpectraST FASTA		E. coli	<i>E. coli</i>	it	42	54,479	32,480	23%
May 21, 2009	ASCII NIST SpectraST FASTA		rat	<i>R. norvegicus</i>	it	25	20,992	15,206	2%

NISTMSQC: Full Analysis of LC-MS/MS data Library/quality metrics

“Performance Metrics for Liquid Chromatography-Tandem Mass Spectrometry Systems in Proteomics Analyses”, *Molecular & Cellular Proteomics*, **9**, 225, 2010

Tandem MS

NIST Tandem Mass Spectral Library 2012



Compounds	7,020
Precursor Ions	15,517
Spectra	123,781

Fragmentation Type	Precursor Ions
Ion Trap	>10,000
Beam Collision Cell (QTOF, QQQ, HCD)	>8,000

Classes:

Metabolites, Drugs, Sugars,
Phospholipids, Peptides,
Surfactants, etc.

Precursors:

$[M+H]^+$, $[M+2H]^{2+}$, $[M-H]^-$,
 $[M+Na]^+$, $[M+NH_4]^+$, $[Cat]^+$, $[An]^-$,
 $[p-H_2O]$, $[p-NH_3]$, etc.

New Software Features:

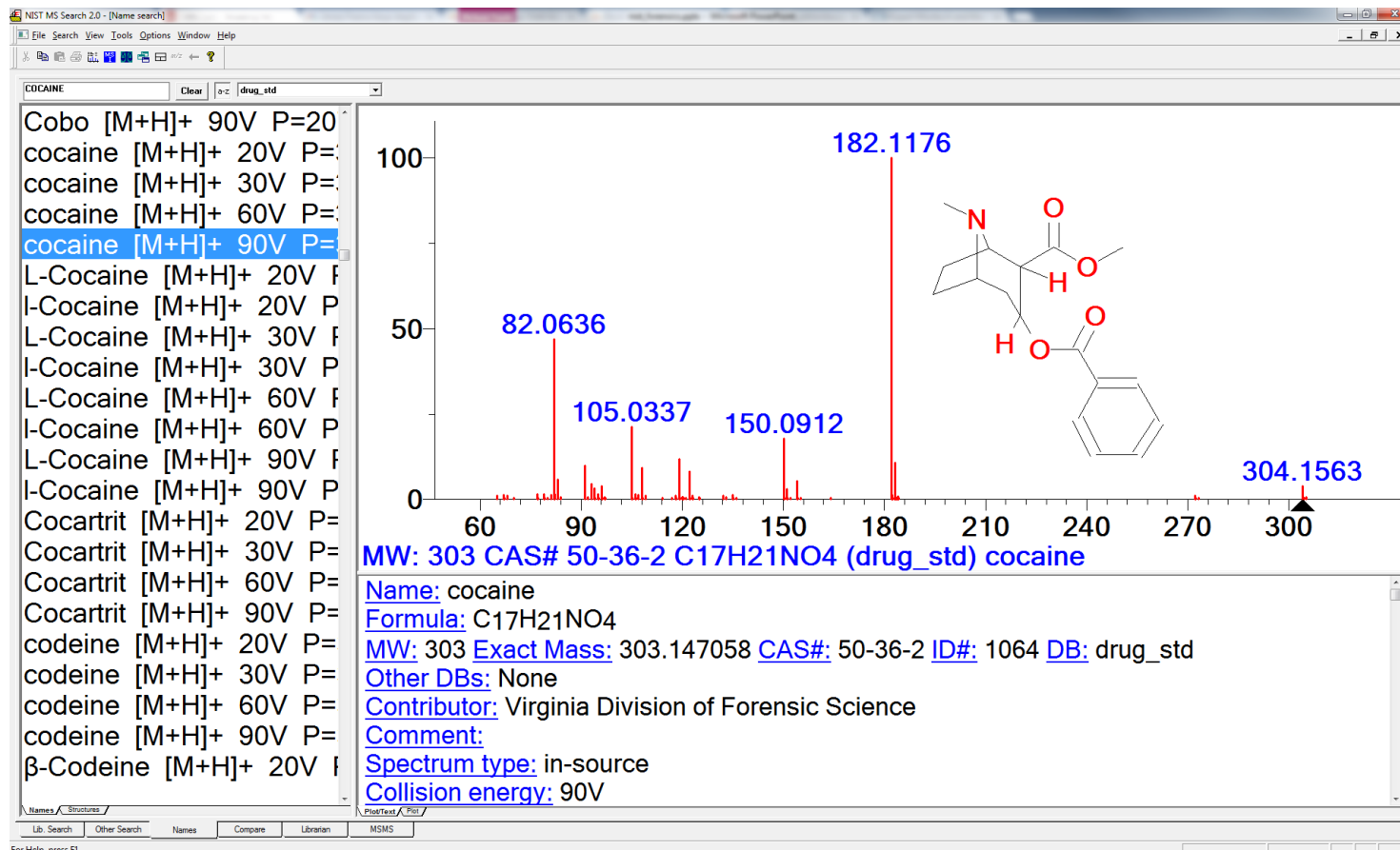
- Exact or isotopic precursor mass & fragment ions.
- Formats: *mzXML*, *mzData*, *mgf*, *msp*, *dta*, *pkl*, *JCAMP*,
- Compatible with NIST EI & Peptide Tandem Libraries.
- New methods for finding targets in the presence of noise.

New Scoring:

- Compounds with few dominant peaks.
- Compensates for *m/z* tuning errors.

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

- 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

SRM + DSRM

- Standard Reference Materials + Data
 - Substance, Thermochemical Data, Chemistry

W



The screenshot shows the SRM/D website interface. At the top, there's a navigation bar with the NIST logo and the text 'SRM/D Standard Reference Material and Data'. Below this, the main content area features a large image of a vial labeled 'STANDARD REFERENCE 1950 Metabolites in Human Plasma' on the left. To the right of the vial is a section titled 'About SRM/D:' which includes a list of bullet points and a small image of a lab technician. The bottom of the page contains the 'MML Material Measurement Laboratory' logo and contact information.

SRM/D
Standard Reference Material and Data

NIST
National Institute of
Standards and Technology

SRM/D 1950
(Metabolites in Human Plasma) Order an SRM

News:

- 04/18/2012
SRM/D 1950 page has
been launched.

About SRM/D:

- SRM/D was designed to support a growing demand for complex biological reference materials such as human plasma (SRM1950)
- SRM/D pages provide interactive access to experimental data produced at NIST and by its customers organized by SRM
- Since NIST can only certify a limited number of analytes per SRM, these pages provide access to additional analyte data
- Analytical methods used to characterize a material depend on the material, select a material from the list above for details

MML
Material Measurement
Laboratory

Last updated: April 18th 2012
Feedback / Comments: Paul Rudnick - paul.rudnick@nist.gov

SRMD.NIST.GOV

• Start

– A

– S

ents

NIST MS Data Center

