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

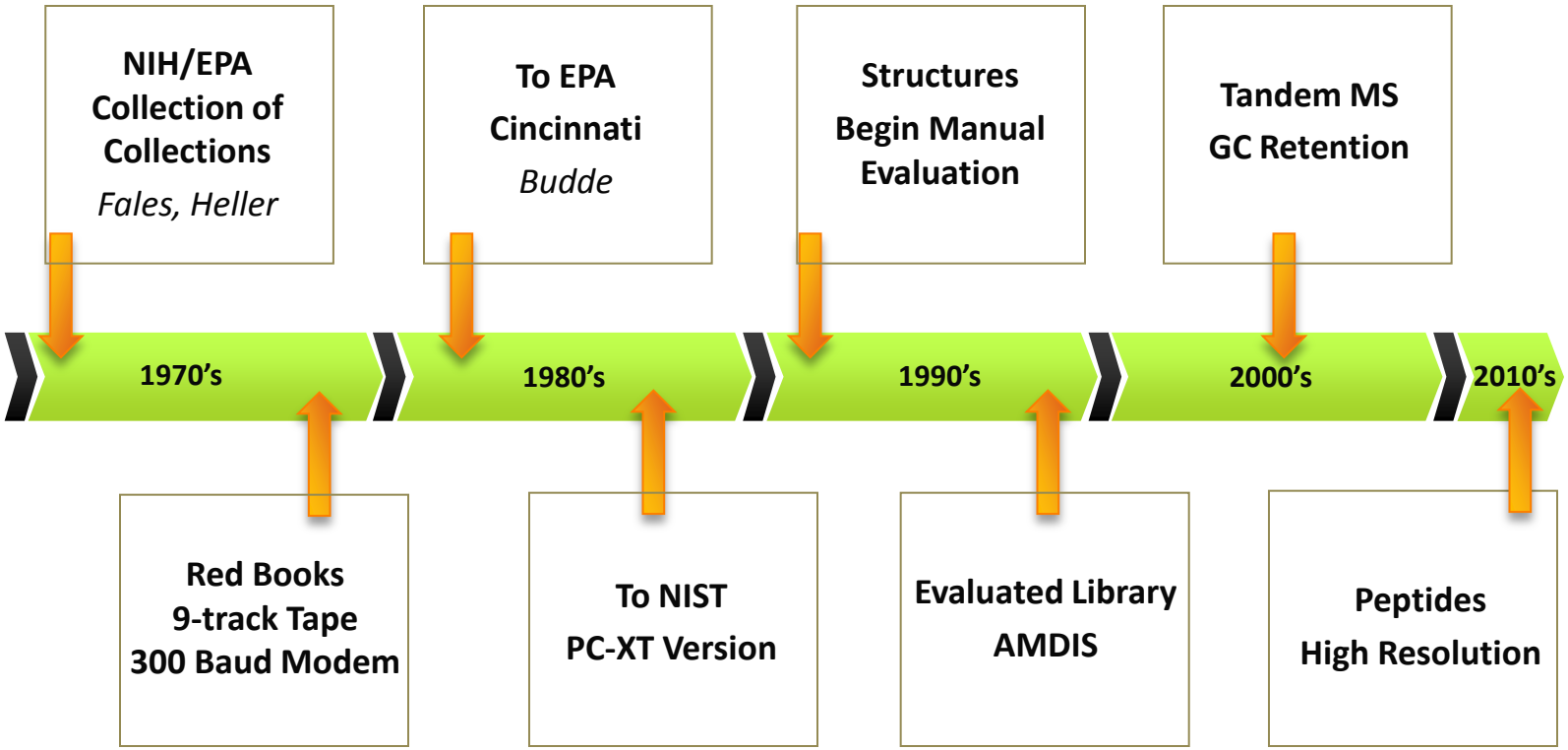
- People
  - DNA, Fingerprints, Features, …
- Objects
  - Clothing, Weapon, …
- Chemicals
  - Molecular Identity

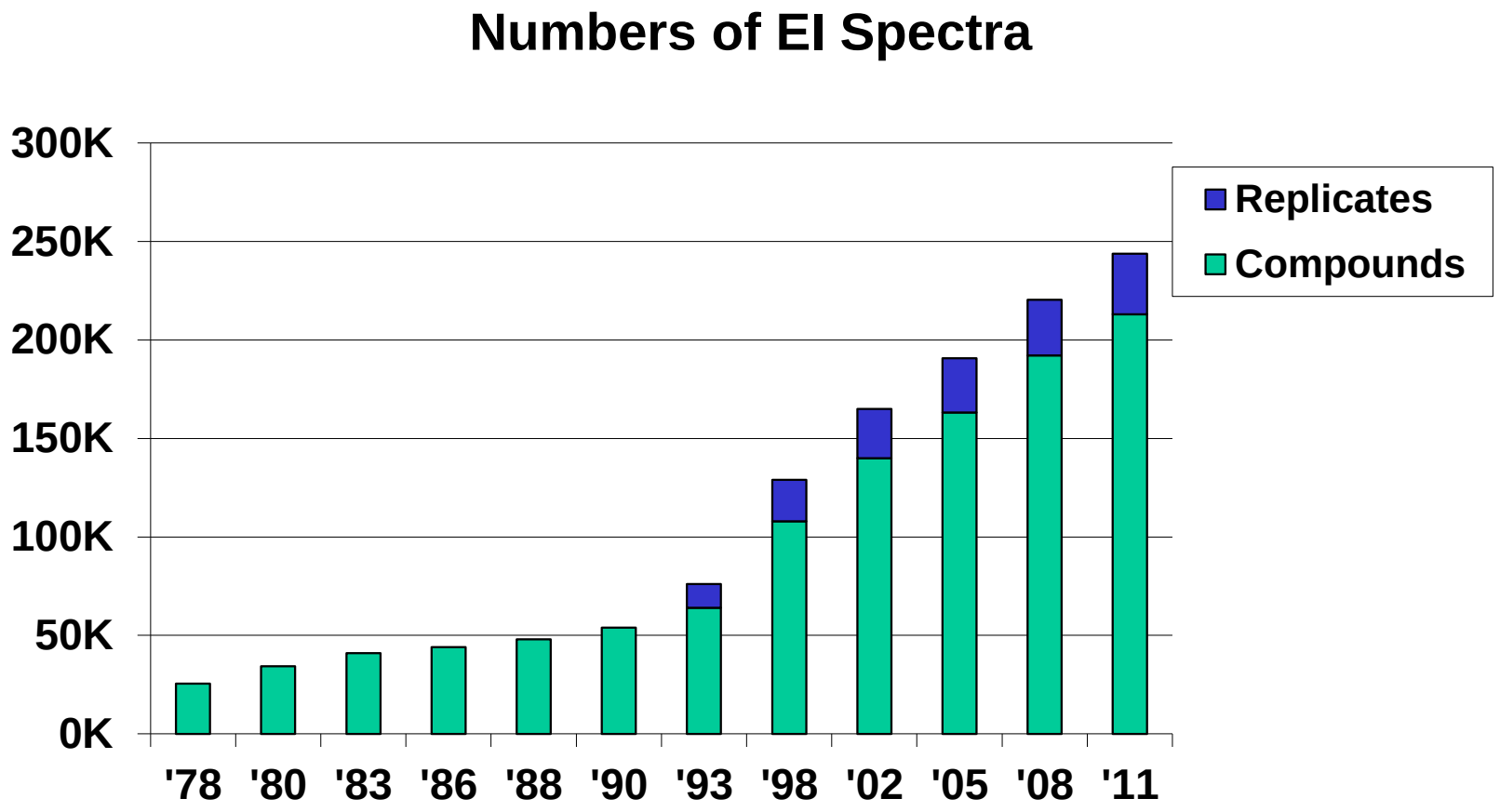
Outline

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

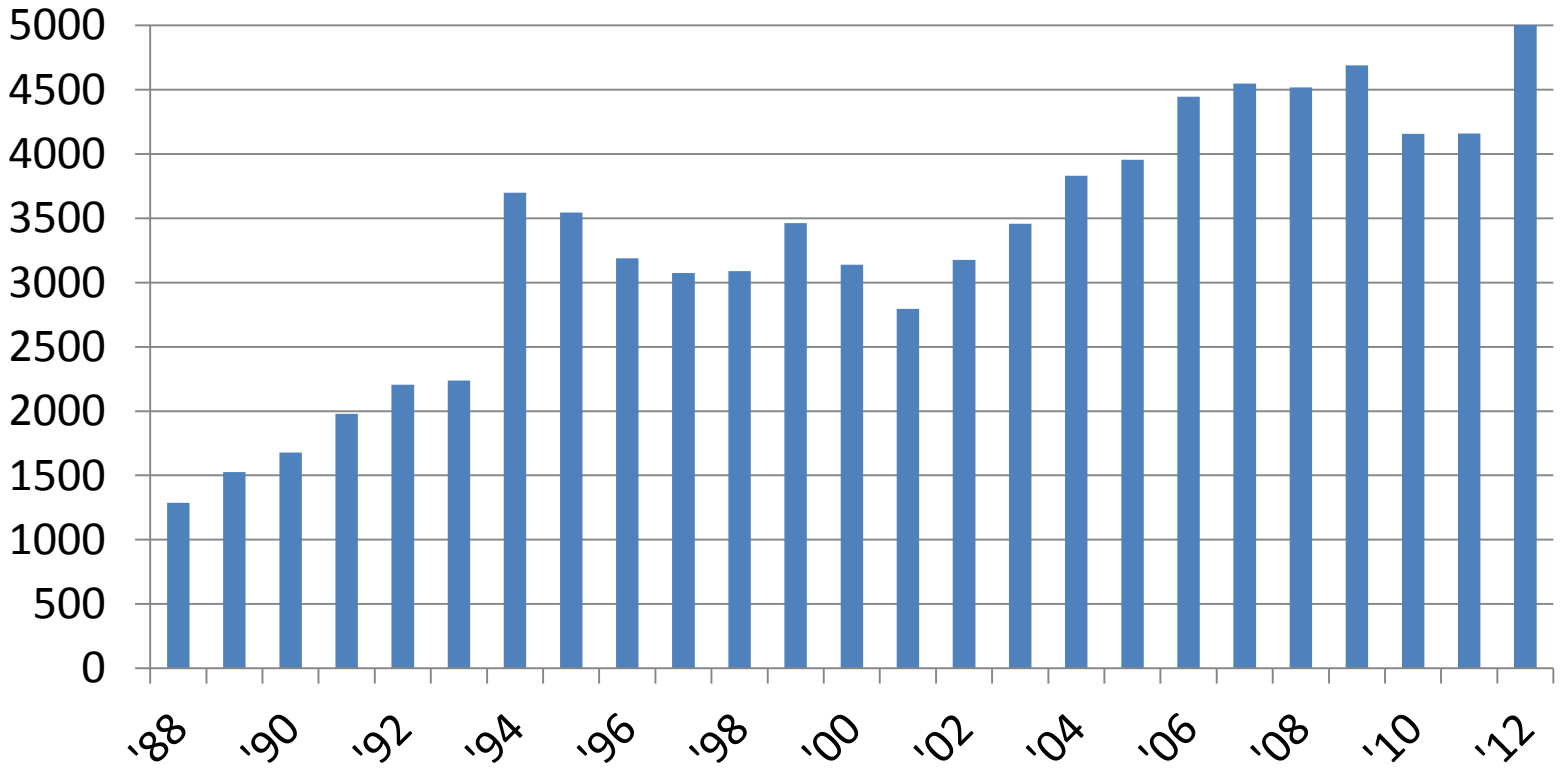
Library Background

Evolution of the NIST MS Library




Numbers of EI Spectra
300K
250K
200K
150K
100K
50K
0K
'78
'80
'83
'86
'88
'90
'93
'98
'02
'05
'08
'11
Replicates
Compounds

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

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 (.hpj)	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

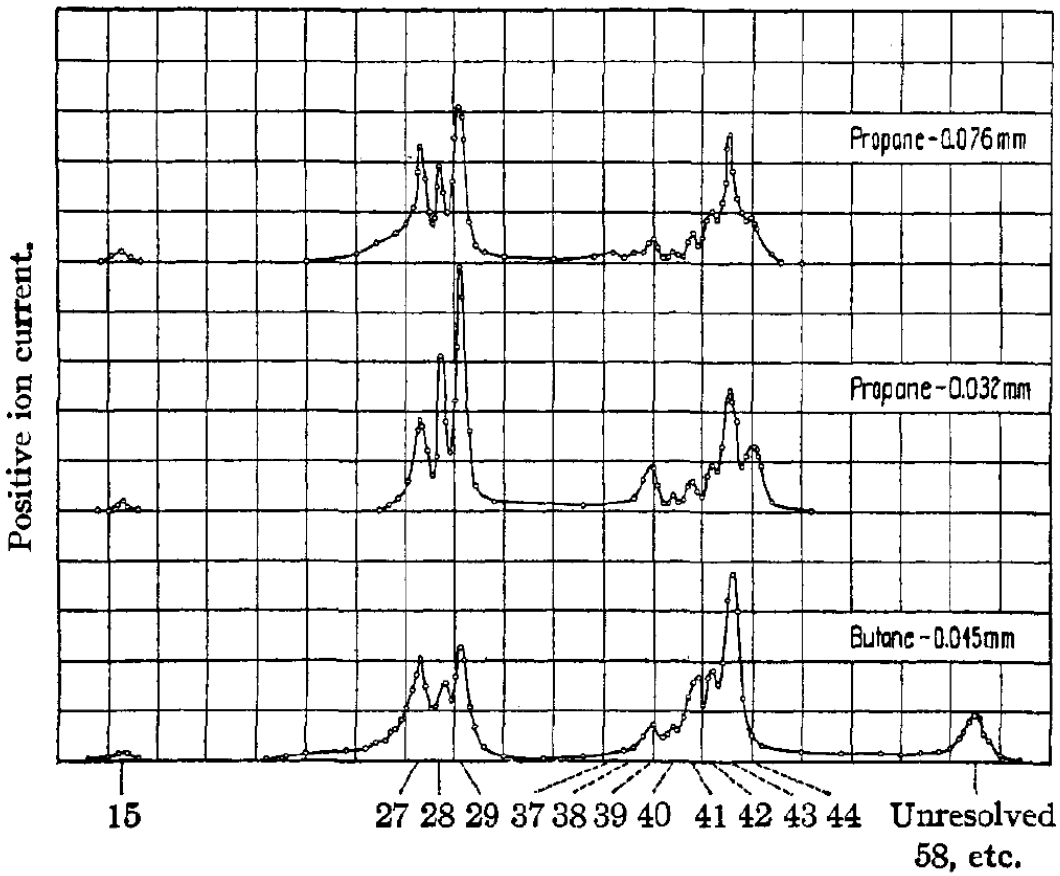
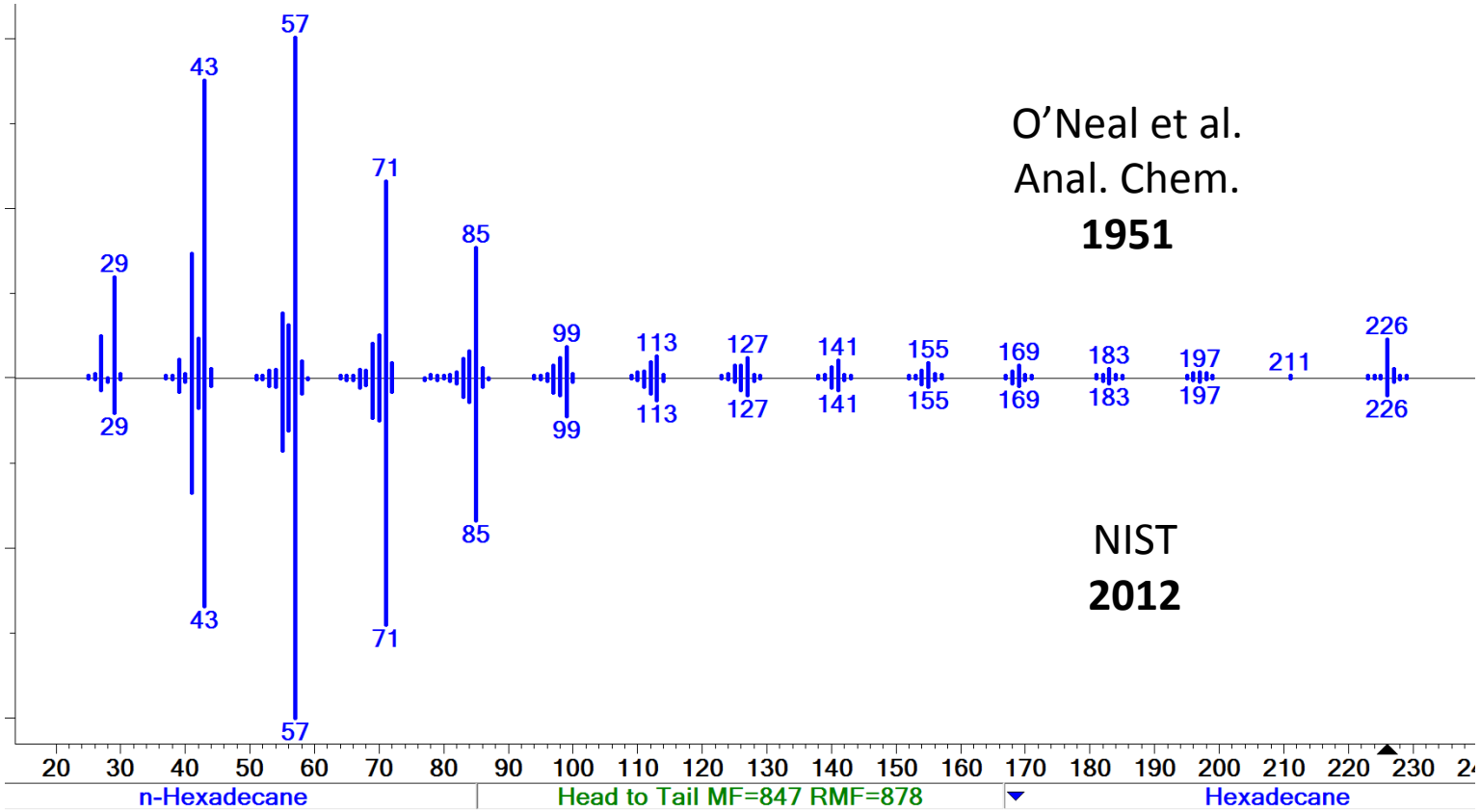


Fig. 3.—Typical runs with propane and butane. Ordinates represent electrometer deflections; abscissas, current through the magnet. Numbers given are calculated masses of corresponding singly charged ions.

JACS, **53**, 1326

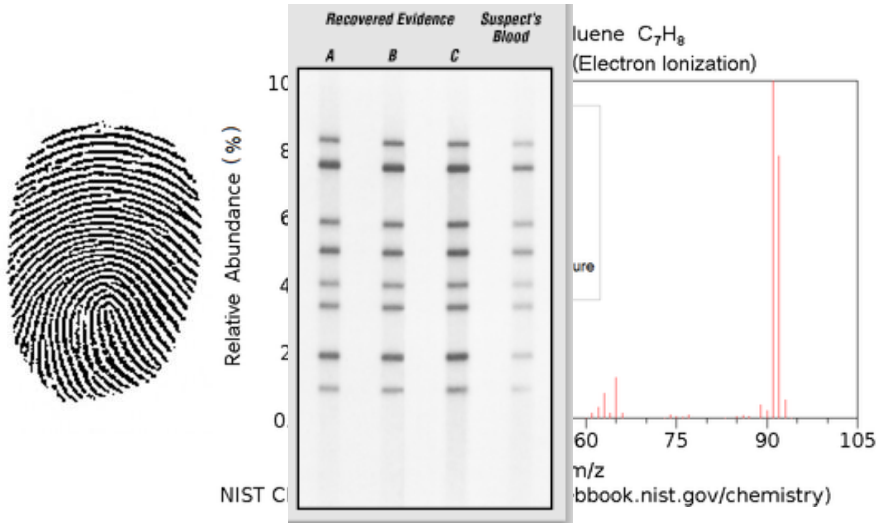
Mass Spectra are Reproducible



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

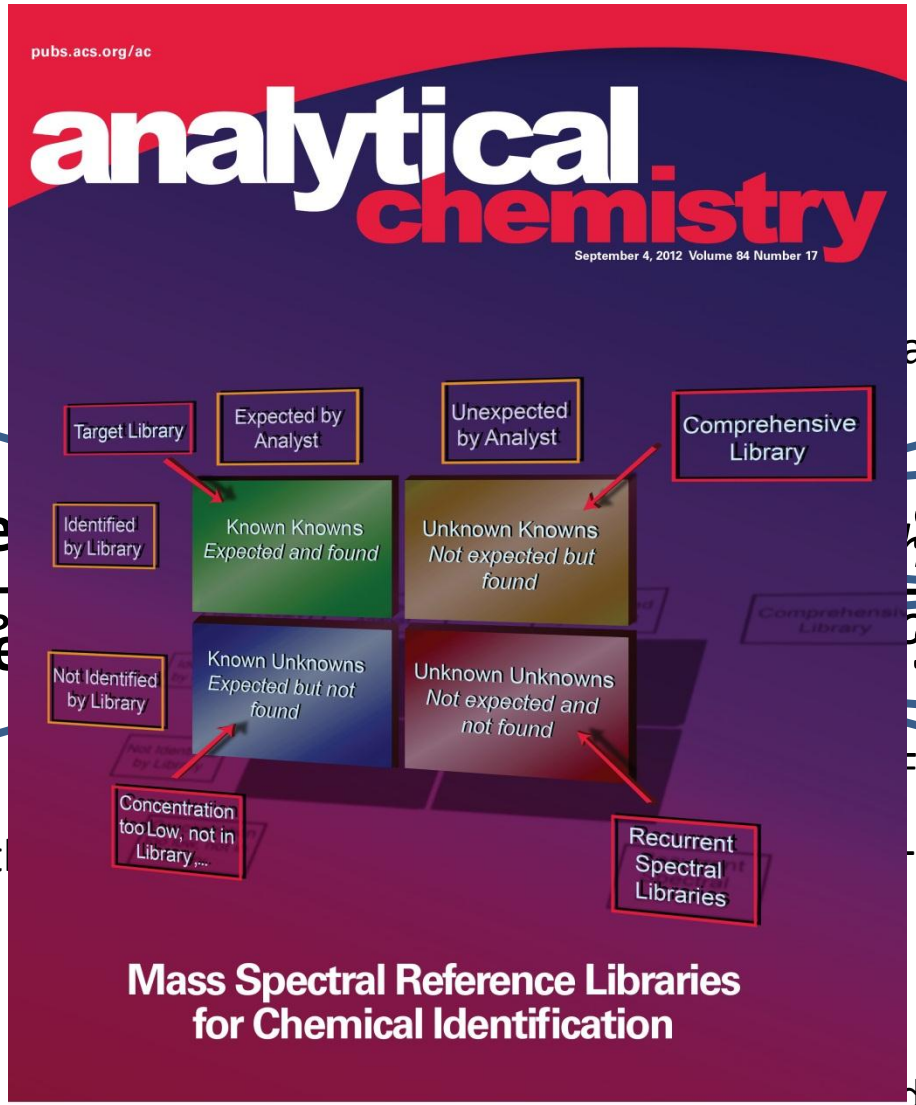
http://en.wikipedia.org/wiki/File:Goldkey_logo_removed.jpg

alse Negative
Potential

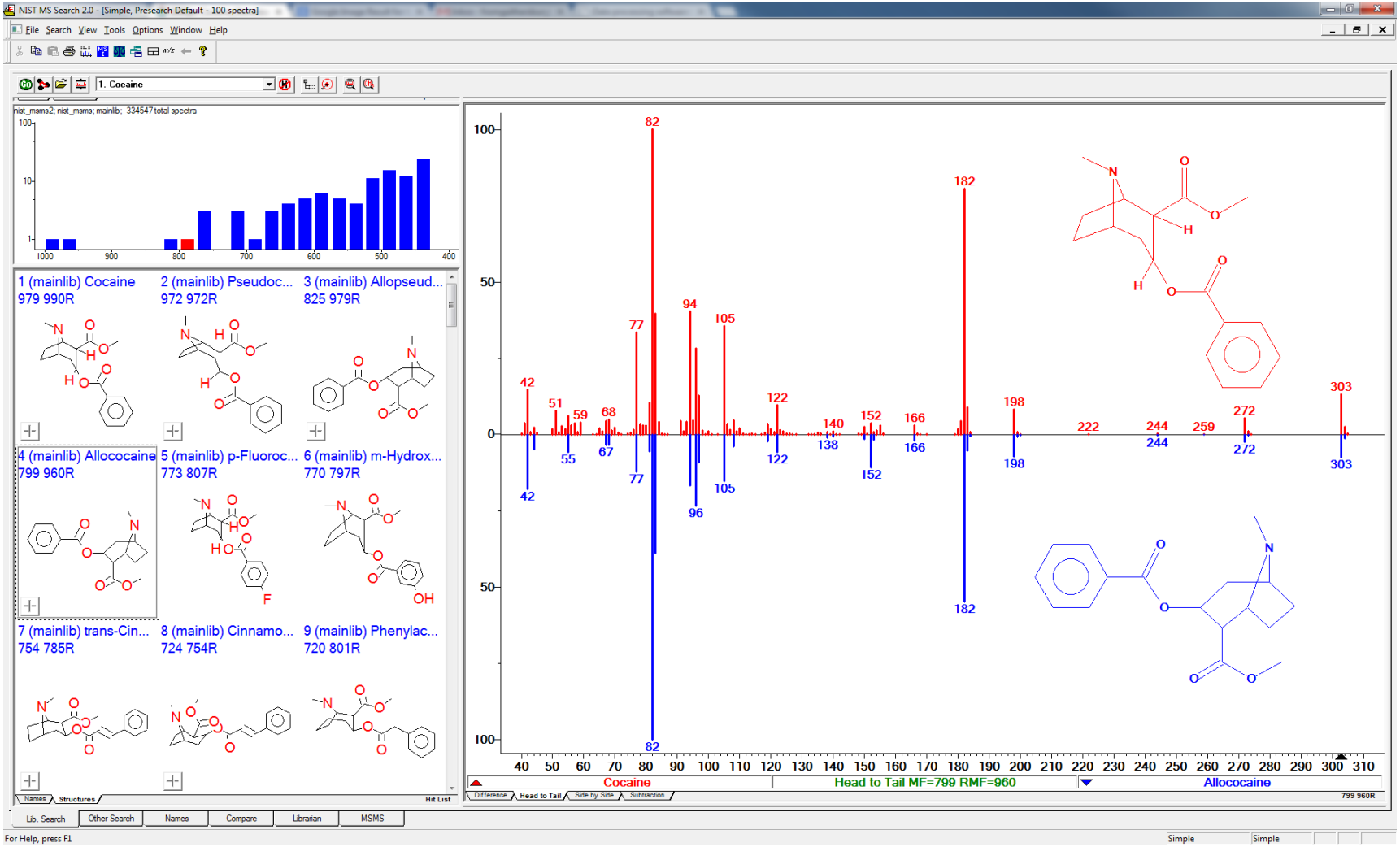
$$\frac{P(ID \mid Score}{P(FP \mid Score} \quad \frac{\text{Change in}}{\text{Confidence}} \quad \frac{Score \mid ID)}{Score \mid FP)}$$

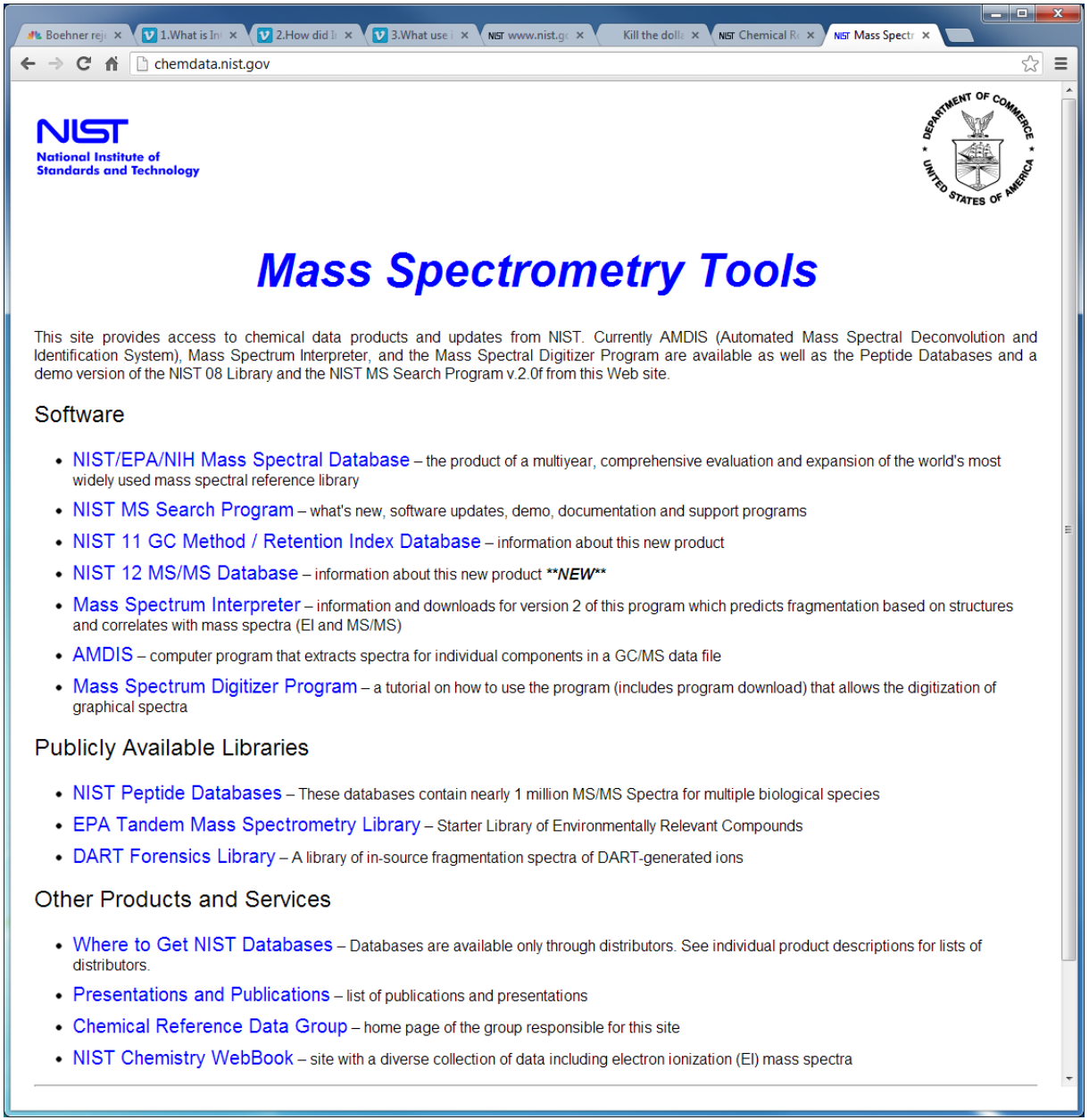
False Positive
Influence of
Potential
Library Search

dds Version



Class Identification or False Positive?





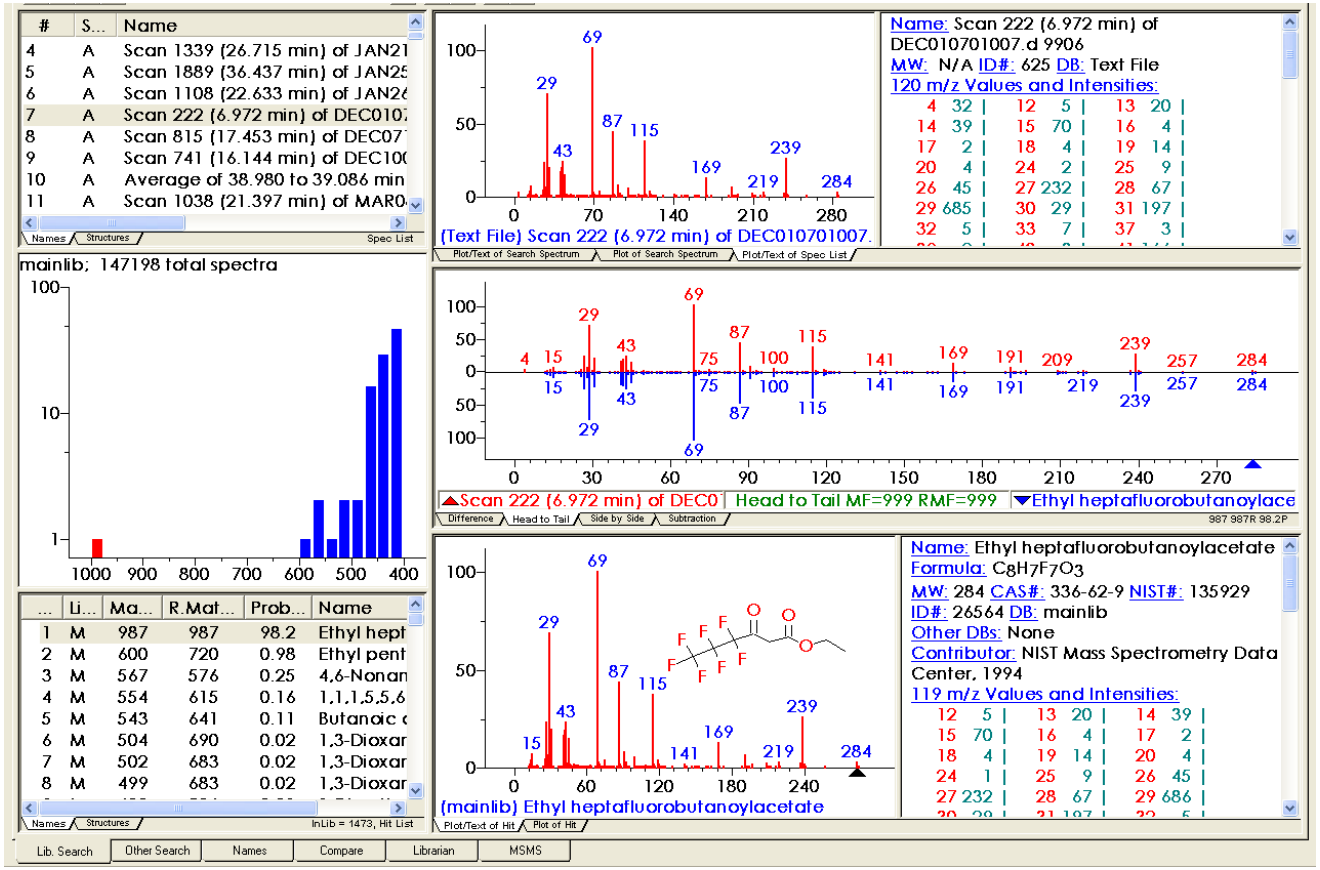
chemdata.nist.gov

Traditional Library Search

Search List

Score Histogram

Hit List

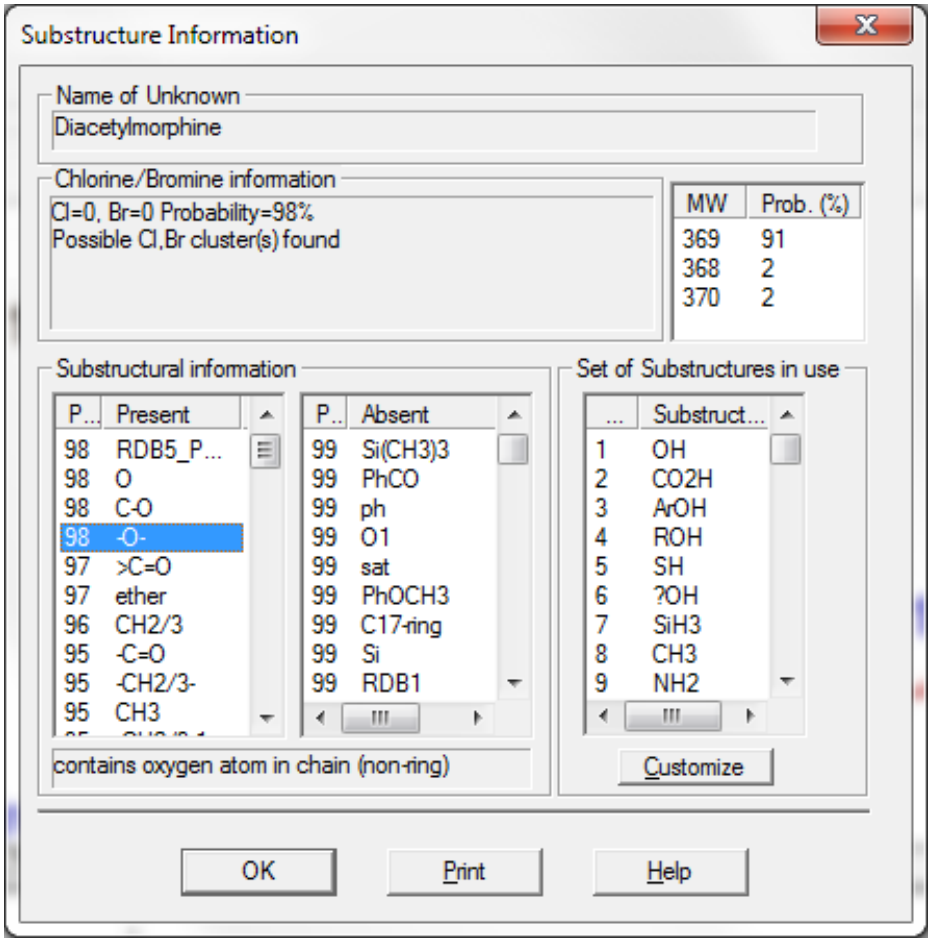


Query Spectrum

Library Spectrum

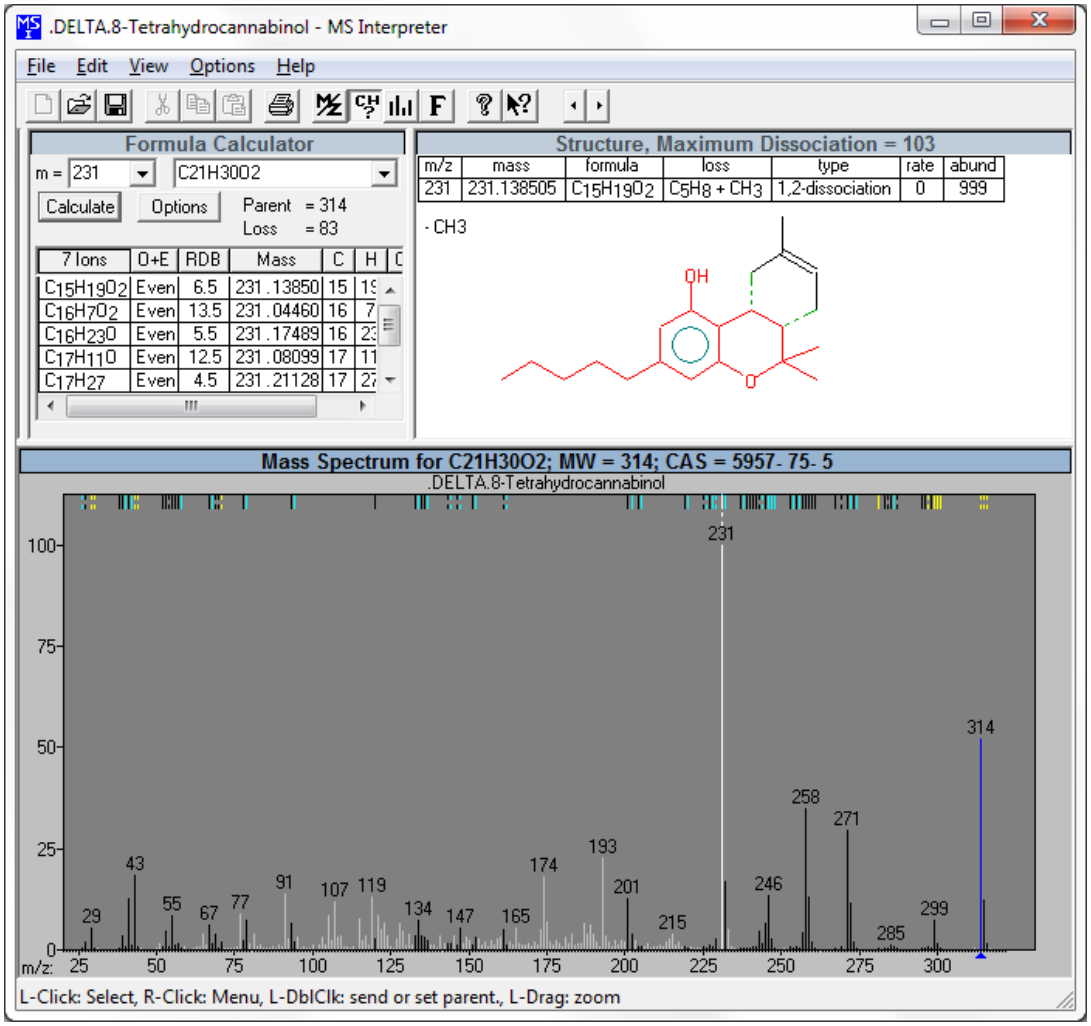
2011 Version - 213K EI, 5K CID, 71K RI Compounds

Substructure Analysis

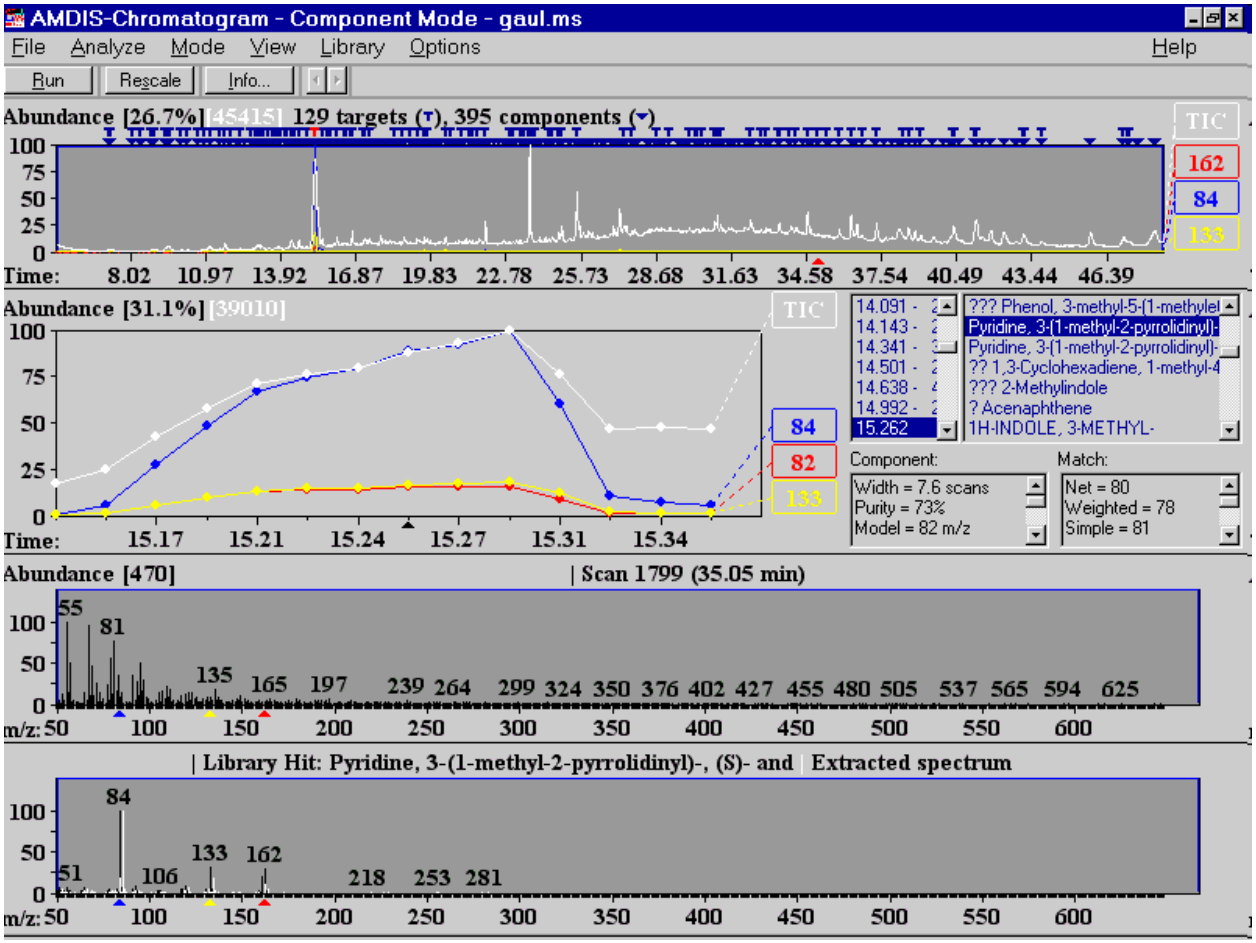


Chemical Substructure Identification by Mass-Spectral Library Searching
JASMS **6** (8) 644-655 (1995)

MS Interpreter



AMDIS



Automated Mass Spectral Deconvolution and Identification System

Created for Chemical Weapons Treaty Verification: “Blinded CW Identification”

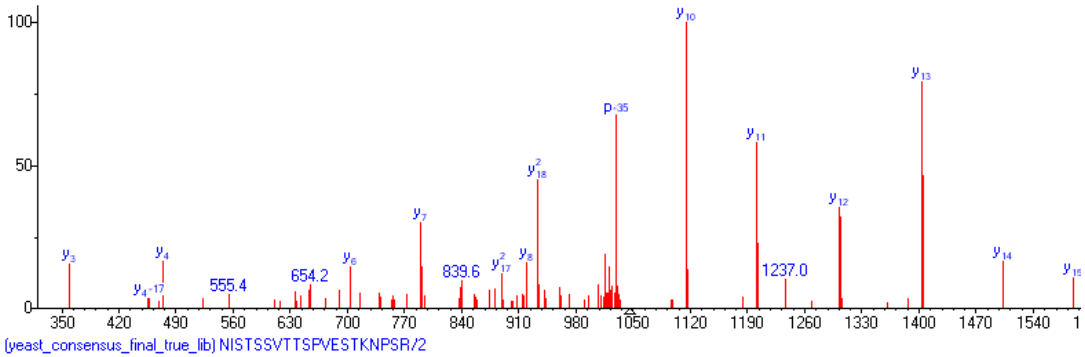
JASMS **1999** *10* 770-781

NIST Scientific and Technical Databases

NIST National Institute of Standards and Technology

Chemical and Biochemical Reference Data

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

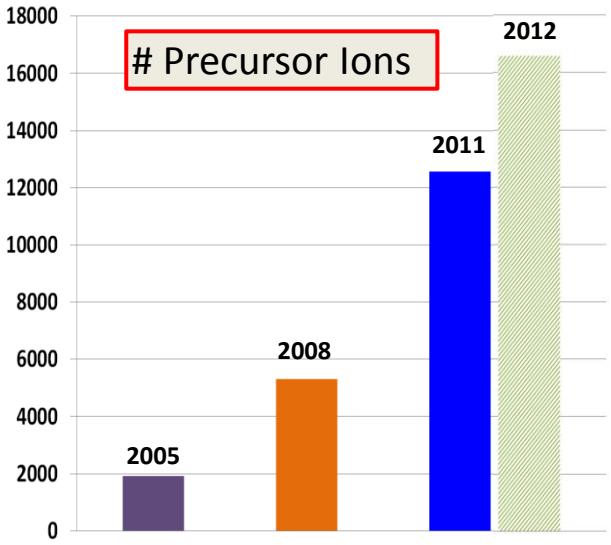
NISTMSQC: Full Analysis of LC-MS/MS data

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.

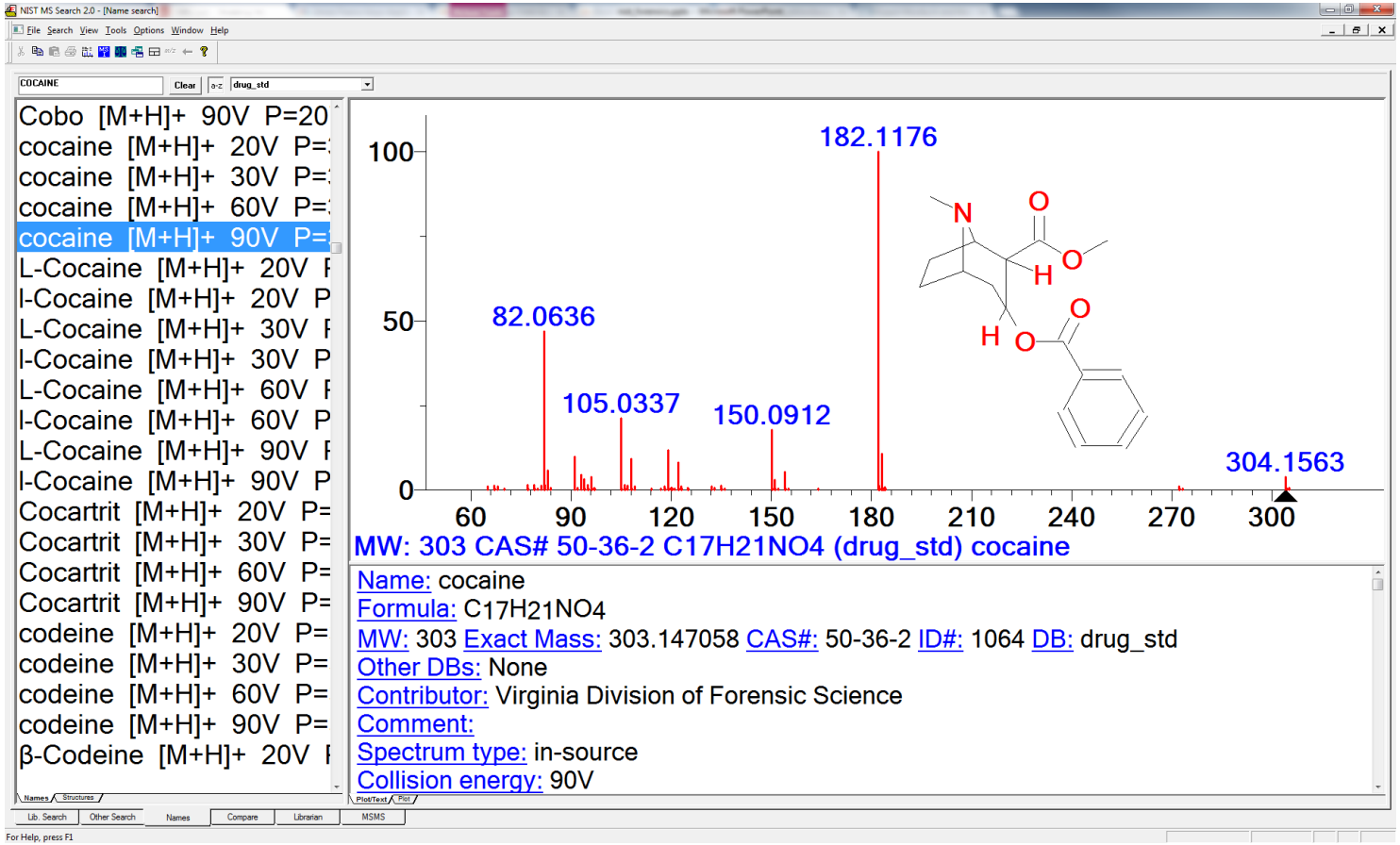
Precursors:
$[M+H]^+$, $[M+2H]^{2+}$, [M-H]-, $[M+Na]^+$, $[M+NH4]^+$, $[Cat]^+$, [An]-, [p-H2O], [p-NH3], etc.

New Software Features:
- Exact or isotopic precursor mass & fragment ions.
- Formats: *mzXML*, *mzData*, *mgf*, *msp*, *dta*, *pkl*, *JCAMP*, ….
- 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

SRD SRM

- Standard Reference Data Materials + Data
  - Substance, Thermodynamics Data, Chemistry W
- Star
  - A
  - Su
    - ents

SRMD.NIST.GOV

NIST MS Data Center