The NIST Mass Spectrometry Data Program

- History
- Research
- NIST 02
  - New Spectra
  - Evaluation
  - Algorithms
  - Distribution
- Pricing & Policies

Distributors' Meeting
PITTCON '02
March 19, 2002

Public Law 90-396
90th Congress, H. R. 6279
July 11, 1968

An Act

To provide for the collection, compilation, critical evaluation, publication, and sale of standard reference data.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

Standard Reference Data Act.

DECLARATION OF POLICY

SECTION 1. The Congress hereby finds and declares that reliable standardized scientific and technical reference data are of vital importance to the progress of the Nation's science and technology. It is therefore the policy of the Congress to make critically evaluated reference data readily available to scientists, engineers, and the general public. It is the purpose of this Act to strengthen and enhance this policy

DEFINITIONS

SEC. 2. For the purposes of this Act—

(a) The term "standard reference data" means quantitative information, related to a measurable physical or chemical property of a substance or system of substances of known composition and structure, which is critically evaluated as to its reliability under section 3 of this Act.

(b) The term "Secretary" means the Secretary of Commerce.

Collection and publication of standard reference data.

SEC. 3. The Secretary is authorized and directed to provide or arrange for the collection, compilation, critical evaluation, publication, and dissemination of standard reference data. In carrying out this program, the Secretary shall, to the maximum extent practicable, utilize the reference data services and facilities of other agencies and instrumentalities of the Federal Government and of State and local governments, persons, firms, institutions, and associations, with their consent and in such a manner as to avoid duplication of those services and facilities. All agencies and instrumentalities of the Federal Government are encouraged to exercise their duties and functions in such manner as will assist in carrying out the purpose of this Act. This section shall be deemed complementary to existing authority, and nothing herein is intended to repeal, supersede, or diminish existing authority or responsibility of any agency or instrumentality of the Federal Government.

Standards, etc. Publication in Federal Register.

SEC. 4. To provide for more effective integration and coordination of standard reference data activities, the Secretary, in consultation with other interested Federal agencies, shall prescribe and publish in the Federal Register such standards, criteria, and procedures for the preparation and publication of standard reference data as may be necessary to carry out the provisions of this Act.

Sale of reference data. Cost recovery.

SEC. 5. Standard reference data conforming to standards established by the Secretary may be made available and sold by the Secretary or by a person or agency designated by him. To the extent practicable and appropriate, the prices established for such data may reflect the cost of collection, compilation, evaluation, publication, and dissemination of the data, including administrative expenses; and the amounts received shall be subject to the Act of March 3, 1901, as amended (15 U.S.C. 271–278e).

31 Stat. 1449; Ante, p. 24

U. S. copyright and renewal rights.

61 Stat. 655; 76 Stat. 446.

SEC. 6. (a) Notwithstanding the limitations contained in section 8 of title 17 of the United States Code, the Secretary may secure copyright and renewal thereof on behalf of the United States as author or proprietor in all or any part of any standard reference data which

MS Library History

1971 – EPA/NIH Collection of Collections

1978 – First Distribution – Tape, On-Line, Books

1983 – To EPA, Cincinnati

1988 – To NIST

1990 – Manual Evaluation/Algorithms

1998 – Evaluated Library

2002 – Major Update

Search Research

Instrument dependence of EI spectra and effects on library searching.

Current Status

- Variety of instruments available for acquisition of mass spectra
  - Magnetic Sector
  - Quadrupole
  - Ion Trap
  - Time of Flight
- NIST Mass Spectral Database
  - Most data comes from quadrupole mass spectrometers

Question: Can I use non-quad data to search the NIST MS library??

Answer:

- "Yes"
  - Magnetic Sector
  - Time of Flight

- "Yes" and "Not Quite"
  - Ion Trap

Ion trap data for many compounds is fine for NIST MS library searches.

EXAMPLE: Pesticides have shown good results

Compound Name	Quadrupole	Ion Trap
Anthracene-d10-	87	87
Chlordane	78	87
Chlorothalonil	95	92
Chlorpropham	95	94
Chrysene-d12	87	79
Dimethoate	83	72
Diphenylamine	99	93
Ethion	76	79
Methidathion	95	82
Methoxychlor	86	61 data is noisy
Permethrin	91	75
o-Phenylphenol	100	96
Phosmet	84	co-elutes with Trifluralin
Pirimiphos methyl	98	90
Terbufos	98	85
Trifluralin	94	90

Matrix sample GC/MS analysis with MS Library match factors between NIST 98 and Quadrupole and Ion Trap data

Area Requiring Attention

Small molecules w/ oxygen, nitrogen, etc.

- Small hetero atom molecules are known to produce mass spectra that are different between quadrupole and ion trap instruments
  - ion intensities
  - product ions
- Library searching is difficult when
  - spectrum is not unique
  - molecular ion is small/nonexistent

Ion Trap Spectra of small molecules (MW 50-225)

Can I use NIST library or not?

"YES"

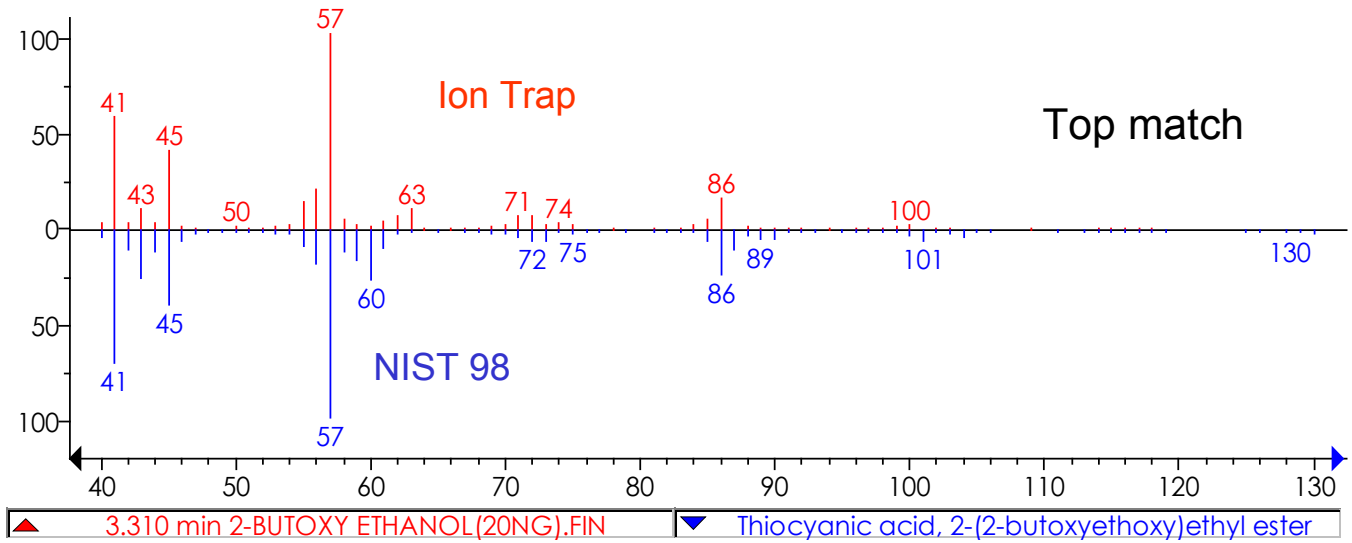
Spectra that produce top matches with NIST 98
Acetic acid, 1-methylethyl ester
Acetic acid, butyl ester
Butanoic acid, 3-oxo-, methyl ester
1-Butanol
1,4-Cyclohexanedimethanol
1,4-Dioxane
Ethanol, 2-mercapto-
Ethanol, 2-phenoxy-
Ethyl Acetate
Glycerin
Hexanal, 2-ethyl-
2-Hexanone, 5-methyl-
Methylene Chloride
N-Methyl-N-(2-hydroxyethyl)carbamic acid, methyl ester
2-Propenoic acid, ethyl ester
Propanoic acid, 3-ethoxy-, ethyl ester
Triethylamine
p-Xylene

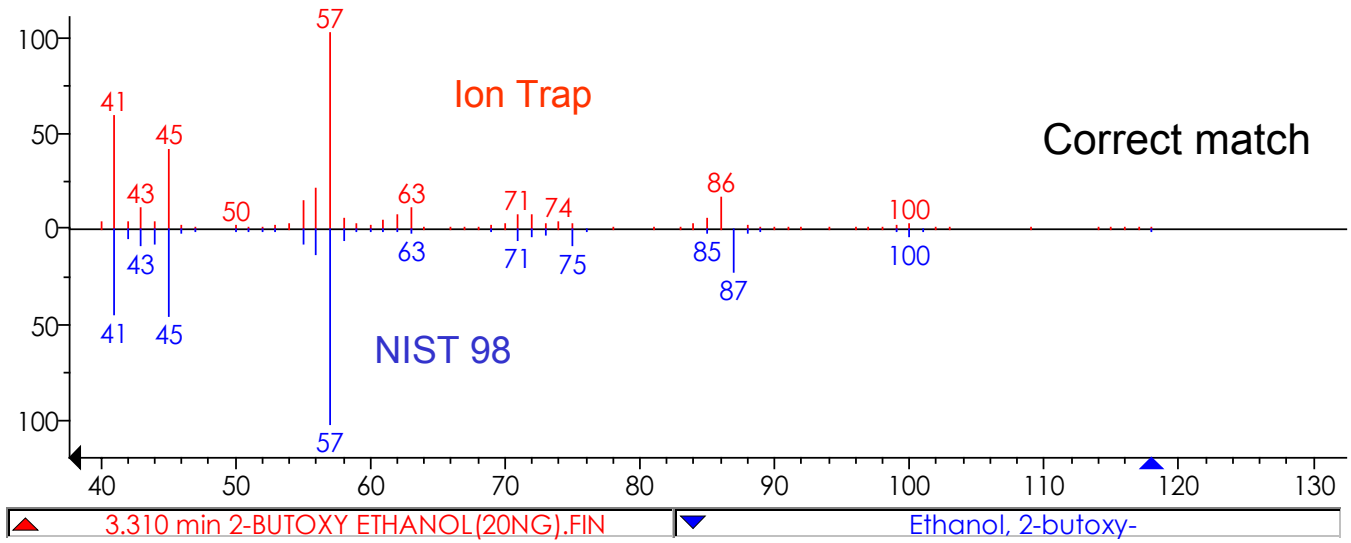
"Not Quite"

within top 10 hits
1-Butanol
2-Butenal
Ethanol, 2-butoxy-
Hexane
Methane, diethoxy-
Pentane
1-Propanol
Propanoic acid, 2-methyl-, 2-methylpropyl ester

incognito in the hit list
1,4-Cyclohexanedimethanol
Ethanol, 2-ethoxy-
Ethyl ether
Hexane
Octane
1-Propanamine

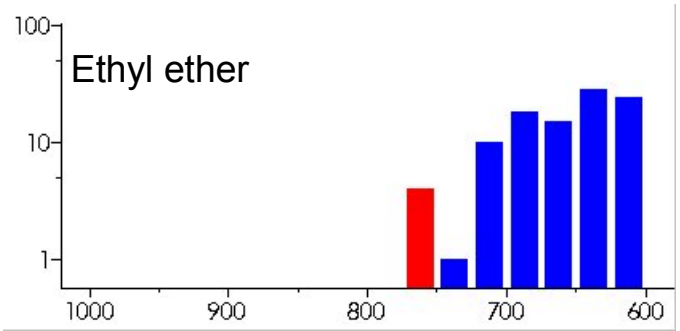
Example

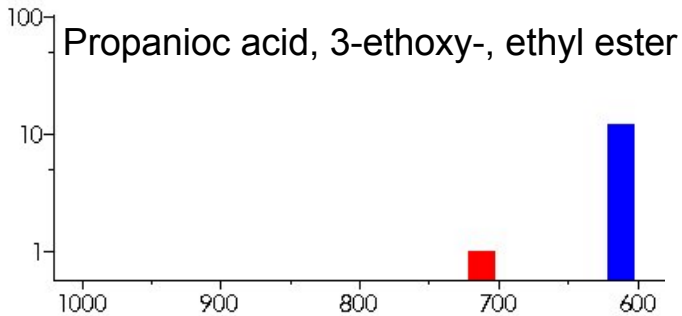

Ion Trap
Top match
NIST 98
3.310 min 2-BUTOXY ETHANOL(20NG).FIN
Thiocyanic acid, 2-(2-butoxyethoxy)ethyl ester


Ion Trap
Correct match
NIST 98
3.310 min 2-BUTOXY ETHANOL(20NG).FIN
Ethanol, 2-butoxy-

Top Library Hit Results

- match factors >700
- probabilities
  - 20-90%(top hit)
  - 2-55%(top 10 list)
  - 1-15% (incognito)





How to Proceed?

- Include ion trap spectra of small molecules in NIST library
  - 60 small molecules
  - 75 pesticides
- Modify library search algorithm to improve search results from ion trap spectra
- Assess variations in spectra from ion trap instruments

MS Data Program

Mass Spectra for Compound Identification

- New Spectra
- Evaluation
- Algorithms
- Distribution

New Spectra

Focus Areas

- New Commercially Available
- Replicates for Important Compounds
- Derivatives
- Chemical Weapons Related

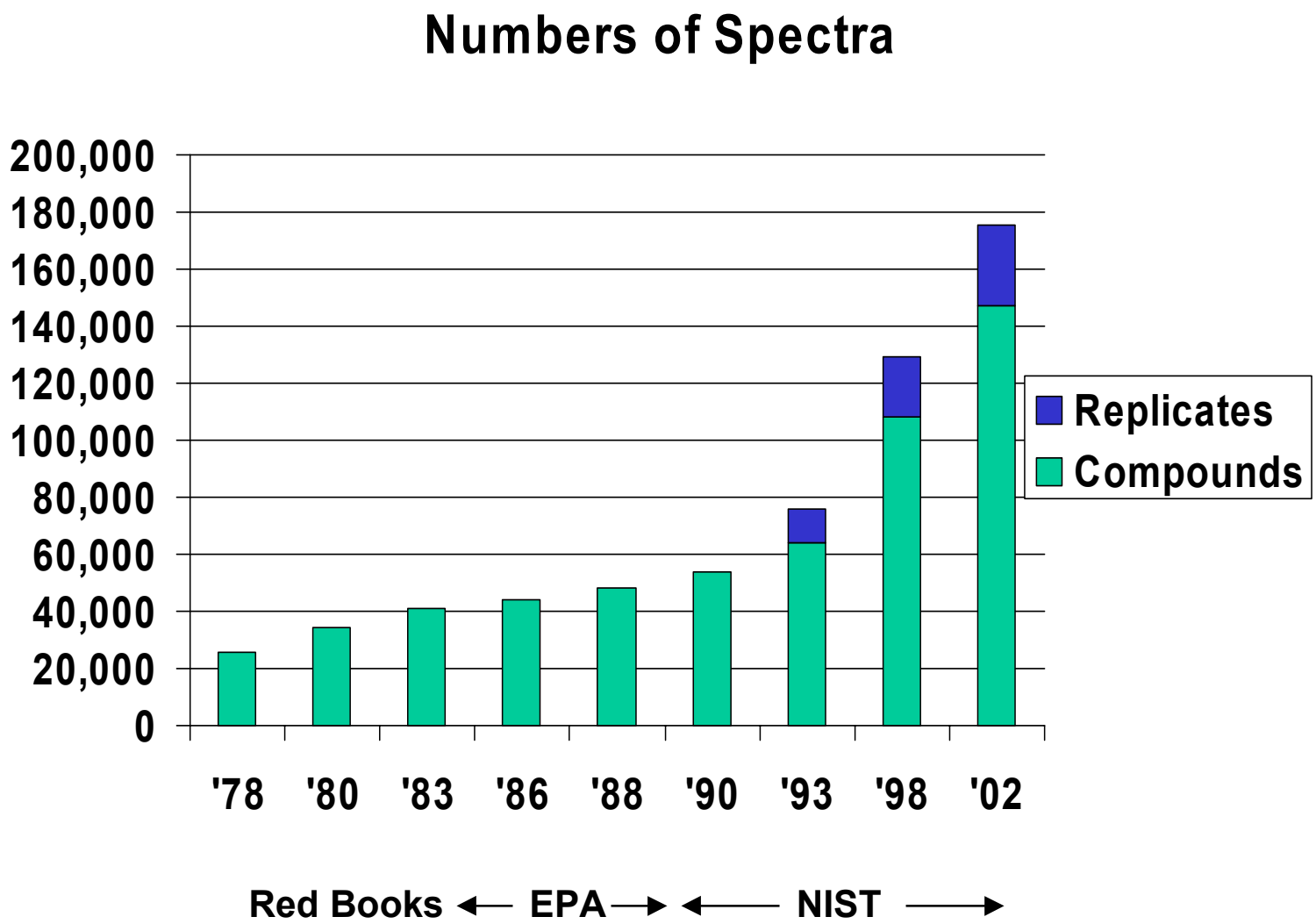
New Spectra

Quality Up Front

- Complete
- With chemical structures
- Documented
- Calibrated instruments
- Upstream filter

Some Sources

- 14,000 Japan AIST/NIMC Collection: Commercially available common organic compounds
- 6,976 Russian Academy of Sciences : Institute of Petrochemical Synthesis:
  - Mostly Derivatives (silyl, acyl).
- 7,182 NIH measurements
  - Synthetic analogues of natural compounds, perspective drugs, drug metabolites, and their intermediates.
- 1,735 NIST
  - Commercially available common compounds, pesticides, drugs.
- 1,022 Eastman Chemical Company
  - Commercial and synthetic compounds and silyl derivatives.
- 406 Verifin (Finland)
  - : Chemical weapons and precursors.
- 348 HD-Science (UK)
  - : Silyl derivatives of drugs
- 138 Military Institute of Chemistry and Radiometry (Poland):
  - Chemical weapons related


Numbers of Spectra
200,000
180,000
160,000
140,000
120,000
100,000
80,000
60,000
40,000
20,000
0
'78
'80
'83
'86
'88
'90
'93
'98
'02
Replicates
Compounds
Red Books
EPA
NIST

Reliability

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, CA

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

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

A description of the methods used to build a high quality, comprehensive reference library of electron-ionization mass spectra is presented. Emphasis is placed on the most challenging part of this project - the improvement of quality by expert evaluation. The methods employed for this task were developed over the course of a spectrum-by-spectrum review of a library containing well over 100,000 spectra. While the effectiveness of this quality improvement task depended critically on the expertise of the evaluators, a number of guidelines are discussed which were found to be effective in performing this onerous and often subjective task. A number of specific examples of the particularly challenging task of spectrum editing are given. (J Am Soc Mass Spectroum 199, 10, 287-299)

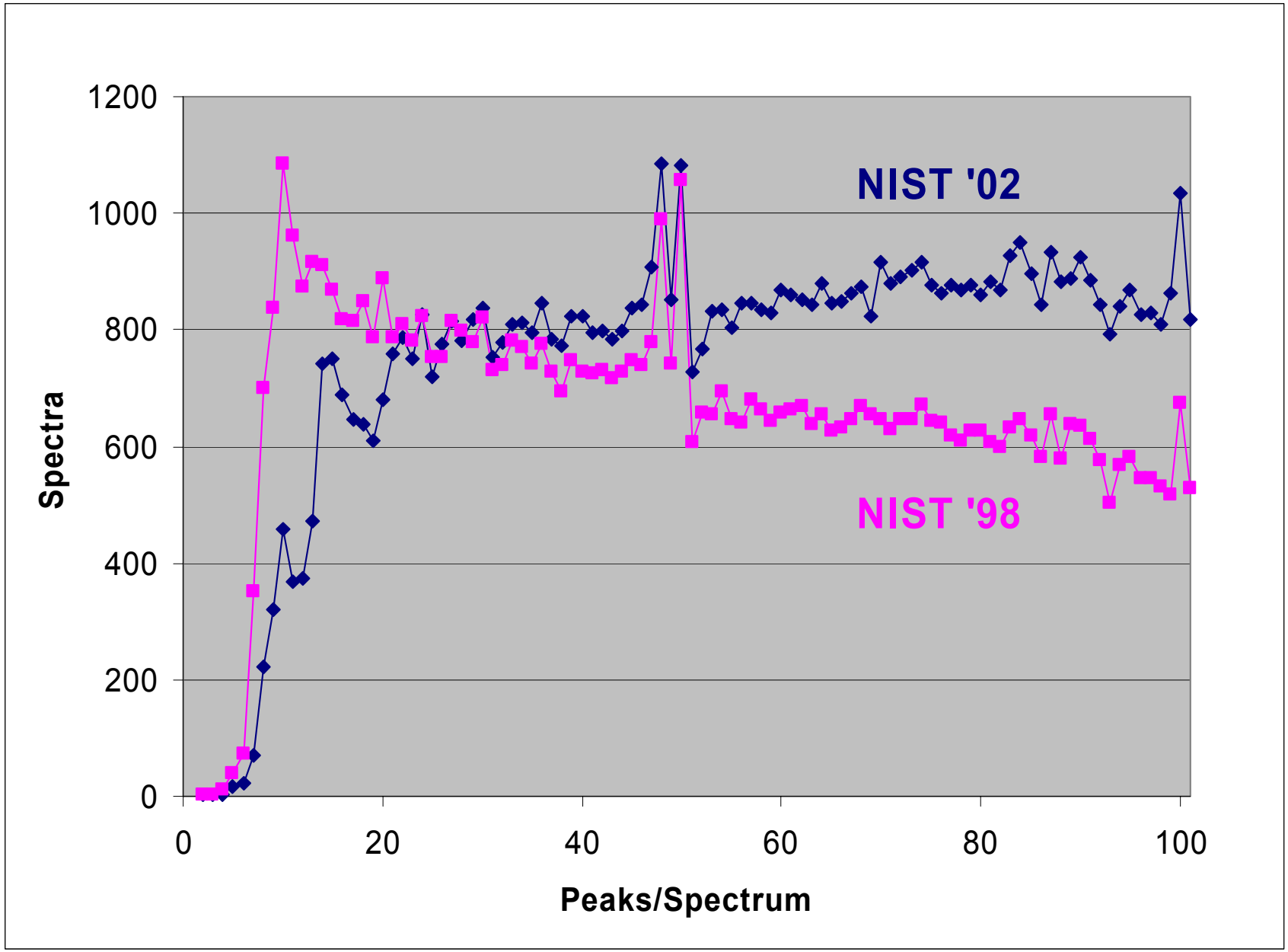
NIST98 => NIST02

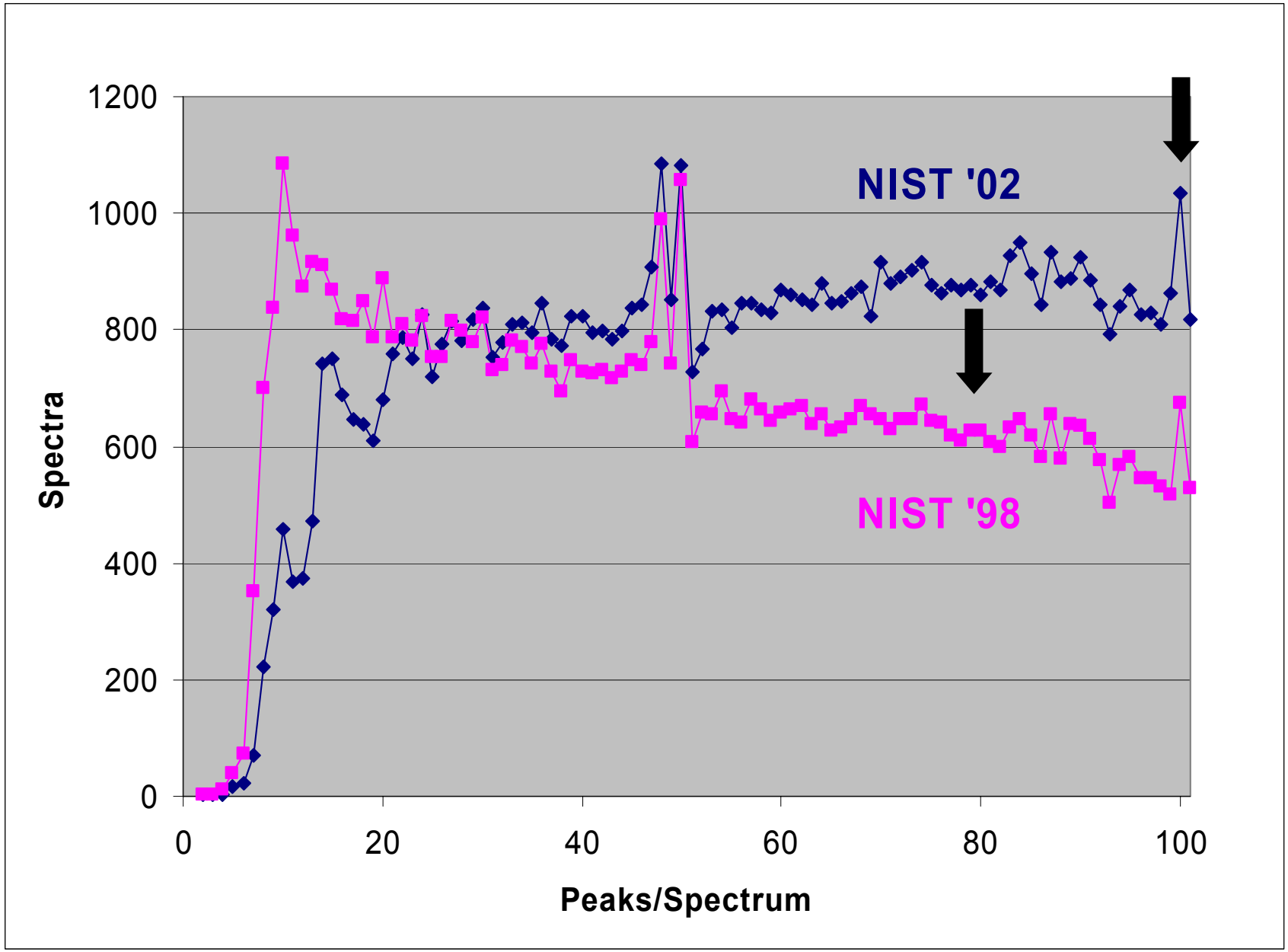
- Mainlib: 107,886 => 147,370 spectra
  - 91,856 spectra from old mainlib
  - 1,331 spectra from old replib
  - 54,183 new spectra
- Replib: 21,250 => 27,844 spectra
  - 14,050 spectra from old replib
  - 7,378 spectra from old mainlib
  - 6,416 new spectra
- Excluded in the new database:
  - 8,652 spectra from old mainlib
  - 5,869 spectra from old replib

More Statistics

- 129,136 => 175,214 spectra
- 90,311 => 134,949 with CAS number
- 69,061 => 107,105 unique CAS numbers
- 107,829 => 147,350 structures
- 255,234 => 440,764 names

- Peaks per Spectrum
  - 79 => 99 median
  - 96 => 111 average
  - 12% => 5% less than 20 peaks/spectrum
  - 2% => 0.5% less then 10 peaks/spectrum


1200
1000
800
600
400
200
0
Spectra
0
20
40
60
80
100
Peaks/Spectrum
NIST '02
NIST '98


1200
1000
800
600
400
200
0
Spectra
0
20
40
60
80
100
Peaks/Spectrum
NIST '02
NIST '98

Quality Control for New Mass Spectra

Compound information data
- ***Formula / Name / Structure / CAS# / MW Consistency***
-

General rules
- ***Confirm Molecular ion and important fragment ions***
- ***Origin of [M-X] ions and the most intense peaks***
- ***Low mass region***
- ***Deduce origin of errors***
-

Compare to other spectra
- ***Replicates, isomers***
- ***Homologous series***
- ***Derivatives***
- ***Spectra in other collections***
-

Spectrum Selection
- ***Completeness***
- ***No Spurious Peaks***
- ***No Saturation***

Select Correct CAS Registry Number for MENTHOL

```
CS: 89781
    E:232390 (-)-MENTHOL
CS: 490993
    E:5285 Cyclohexanol, 5-methyl-2-(1-methylethyl)-,
(1.alpha.,2.beta.,5.beta.)-
 CS: 491010
    E:38796 Cyclohexanol, 5-methyl-2-(1-methylethyl)-,
(1.alpha.,2.alpha.,5.beta.)-
    E:151875 (+)-NEOMENTHOL
CS: 491021
    E:38506 Cyclohexanol, 5-methyl-2-(1-methylethyl)-,
(1.alpha.,2.alpha.,5.alpha.)-
    E:109300 (+)-NEOISOMENTHOL
CS: 1490046
    E:221057 2-Isopropyl-5-methylcyclohexanol
    +E:245179 MENTHOL
CS: 2216515
    E:151883 (-)-MENTHOL
CB: 2216526
    E:108354 (1S, 2S, 5R)-(+)-NEOMENTHOL
CB: 3623516
    C:3623516 Cyclohexanol, 5-methyl-2-(1-methylethyl)-,
(1.alpha.,2.alpha.,5.beta.)-(.+/-.)-
CB: 3623527
    C:3623527 Cyclohexanol, 5-methyl-2-(1-methylethyl)-,
(1.alpha.,2.beta.,5.beta.)-(.+/-.)-
CB: 3623538
    C:3623538 Cyclohexanol, 5-methyl-2-(1-methylethyl)-,
(1.alpha.,2.alpha.,5.alpha.)-(.+/-.)-
CB: 15356602
    E:151874 (+)-MENTHOL
CS: 15356704
    E:290746 DL-Menthol
CB: 20747493
    C:20747493 Cyclohexanol, 5-methyl-2-(1-methylethyl)-, [1R-
(1.alpha.,2.alpha.,5.beta.)]-
CB: 20752345
    C:20752345 Cyclohexanol, 5-methyl-2-(1-methylethyl)-, [1R-
(1.alpha.,2.alpha.,5.alpha.)]-
CB: 23283978
    E:108362 (1S,2R,5R)-(+)-ISOMENTHOL
    E:152083 (+)-ISOMENTHOL
```

5 CAS Registry Numbers for COCAINE Structure

CB: **50362**
E:74557 8-Azabicyclo[3.2.1]octane-2-carboxylic acid, 3-(benzoyloxy)-8-methyl-, methyl ester, [1R-(exo,exo)]-
E:113834 Cocaine
E:246803 COCAINE
E:246808 COCAINE

CB: **478739**
E:120196 PSEUDOCOCAINE
E:248638 PSEUDOCOCAINE

CB: **518978**
E:121444 ALLOPSEUDOCOCAINE

CB: **668199**
E:122751 ALLOCOCAINE

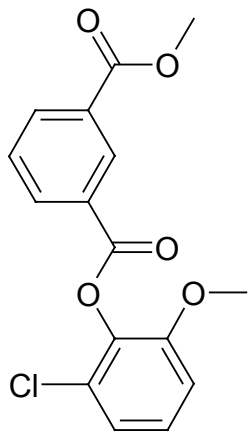
CB: **47195073**
C:47195073 8-Azabicyclo[3.2.1]octane-2-carboxylic acid, 3-(benzoyloxy)-8-methyl-, methyl ester, [1S-(exo,exo)]-

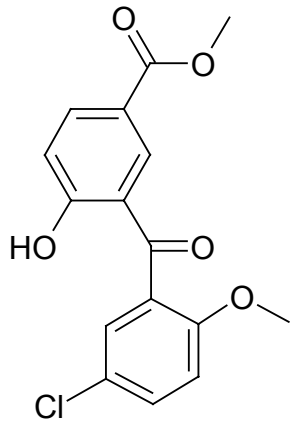
CB: -128222
E:128222 DIHYDROTROPINE, 2-METHOXYFORMYL-3-O-BENZOYL-

CB: -248080
E:248080 PSEUDOALLOCOCAINE

Structure changed to match name

*Benzoic acid, 3-(**5**-chloro-2-methoxybenzoyl)-**4**-hydroxy-*

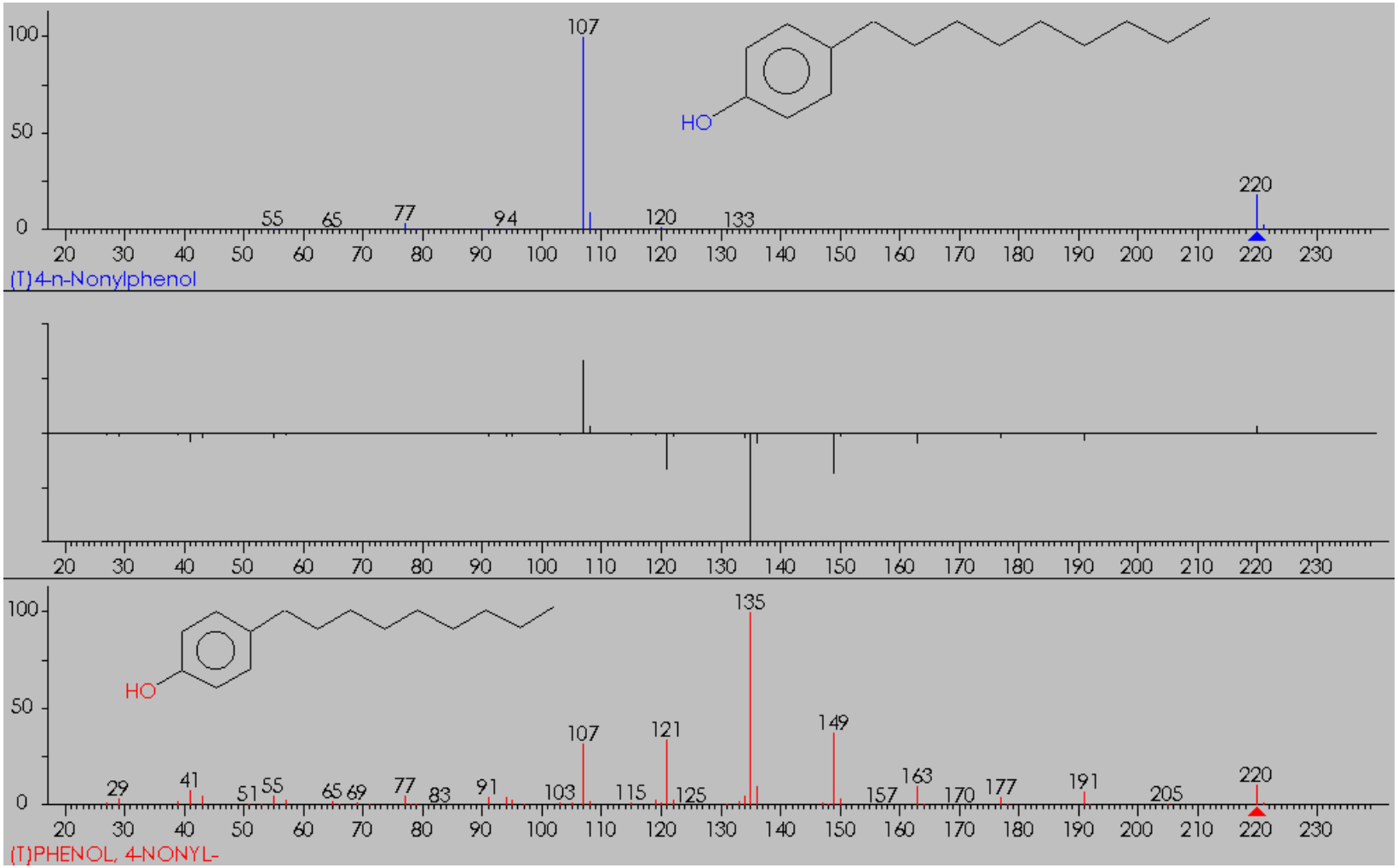


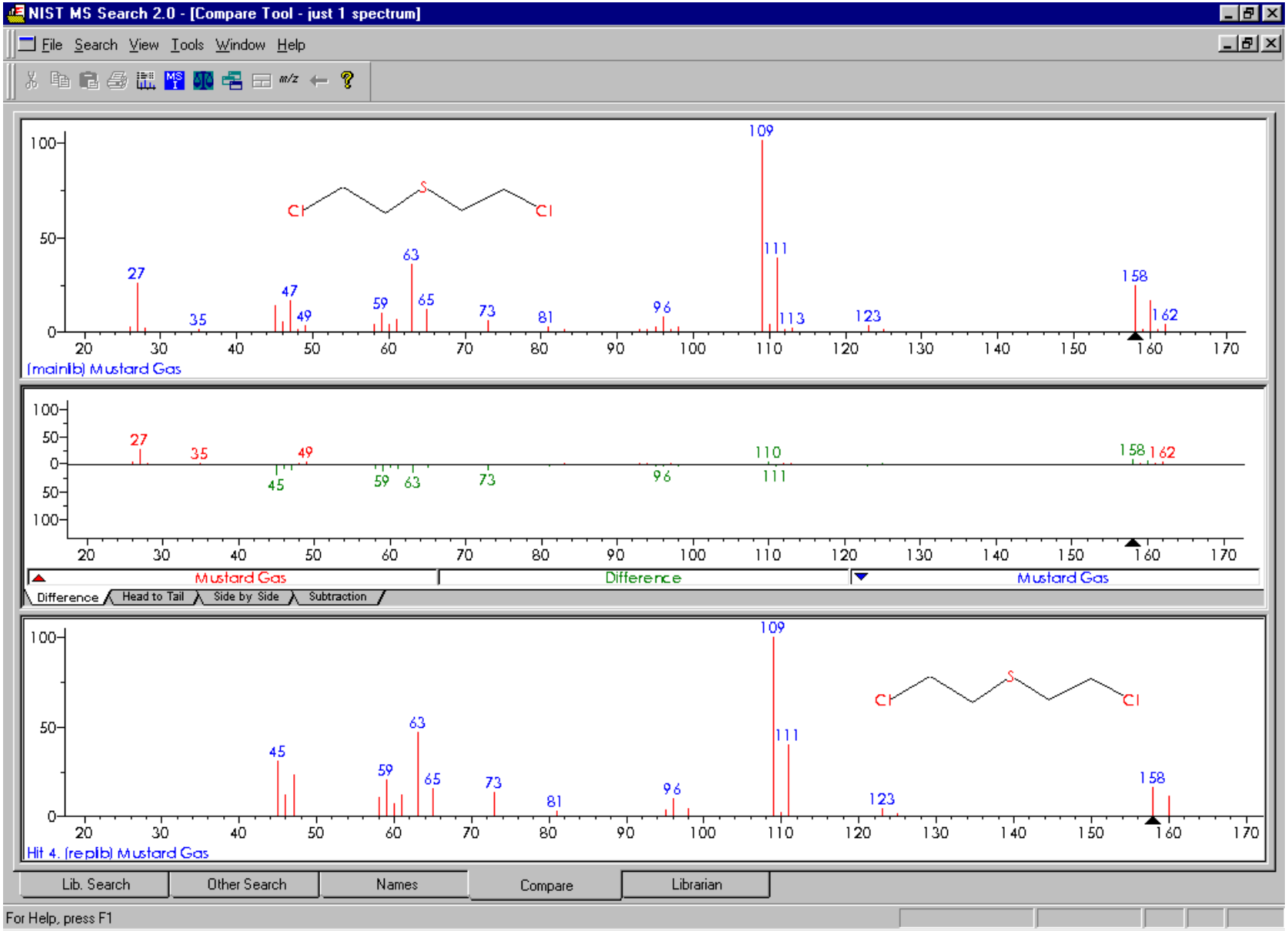


19 Spectra for ASPIRIN

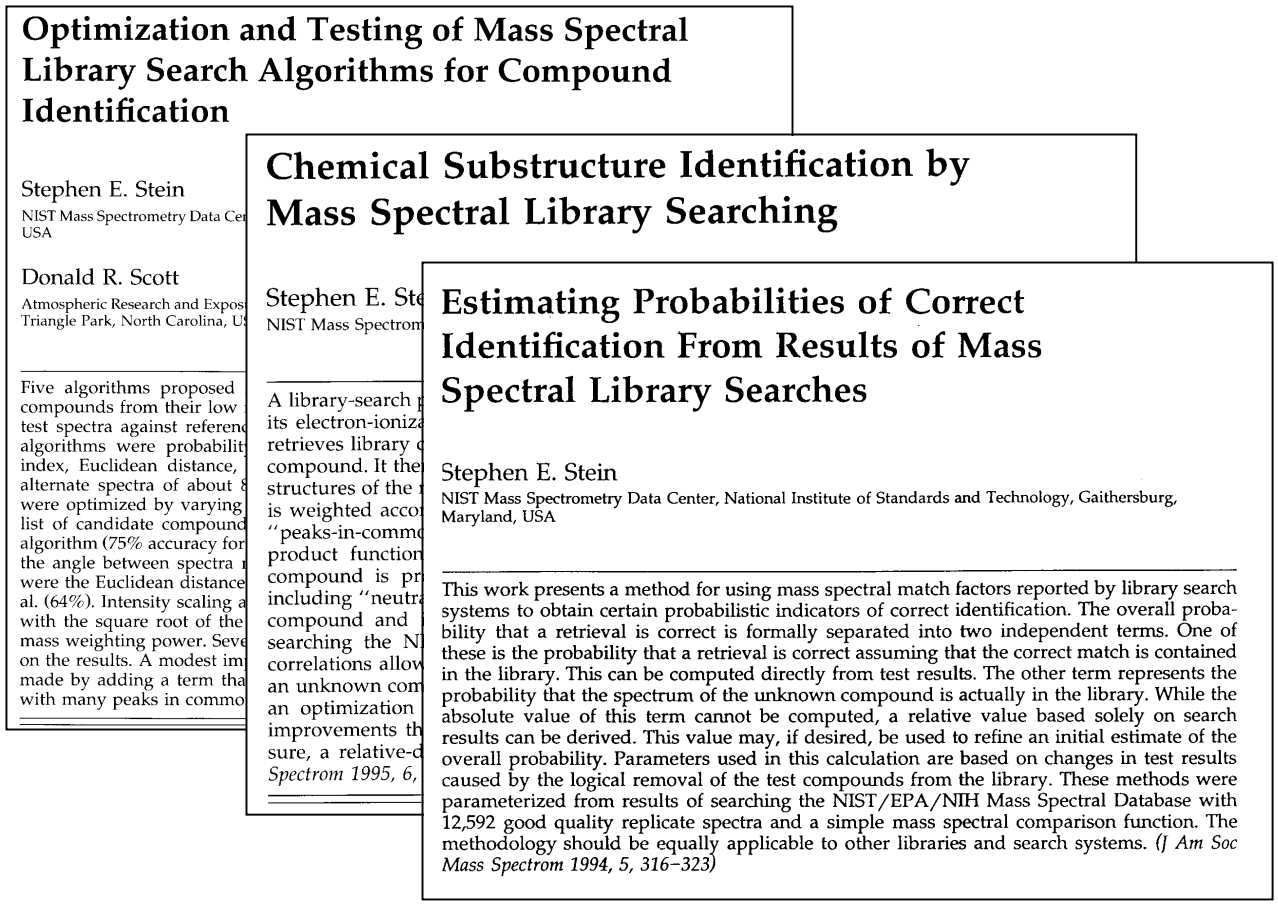
Benzoic acid, 2-(acetyloxy)-	25407	“7”
.Benzoic acid, 2-(acetyloxy)-	23813	“7”
.Benzoic acid, 2-(acetyloxy)-	79540	“7”
.Benzoic acid, 2-(acetyloxy)-	22205	“7”
.Benzoic acid, 2-(acetyloxy)-	157126	“7”
.Acetylsalicylic acid	119370	“7”
.Benzoic acid, 2-(acetyloxy)-	189933	“7”
.Benzoic acid, 2-(acetyloxy)-	126014	“7”
.2-Acetoxybenzoic acid	221215	“Sel”
.o-Acetoxybenzoic acid	227751	“7”
.Acetylsalicylic acid	246070	“7”
.Acetylsalicylic acid	246071	“7”
.Acetylsalicylic acid	246073	“7”
.Acetylsalicylic acid	246079	“8”
.Aspirin	247685	“7”
.Aspirin	250572	“R”
.2- Acetylsalicylic acid	257699	“7”
.2-(Acetyloxy)benzoic acid	287701	“7”
Acetylsalicylic acid	290983	“R”

Replace Mixture Spectrum with “Pure” Spectrum




NIST MS Search 2.0 - [Compare Tool - just 1 spectrum]
File Search View Tools Window Help
(main lib) Mustard Gas
Mustard Gas
Difference
Mustard Gas
Difference
Head to Tail
Side by Side
Subtraction
Hit 4. (replib) Mustard Gas
Lib. Search
Other Search
Names
Compare
Librarian
For Help, press F1

Algorithms:
Search and Error Discovery



Optimization and Testing of Mass Spectral Library Search Algorithms for Compound Identification

Stephen E. Stein

NIST Mass Spectrometry Data Ce
USA

Donald R. Scott

Atmospheric Research and Expos
Triangle Park, North Carolina, U

Five algorithms proposed
compounds from their low
test spectra against referen
algorithms were probabilit
index, Euclidean distance,
alternate spectra of about 8
were optimized by varying
list of candidate compound
algorithm (75% accuracy for
the angle between spectra
were the Euclidean distance
al. (64%). Intensity scaling a
with the square root of the
mass weighting power. Seve
on the results. A modest im
made by adding a term tha
with many peaks in commo

Chemical Substructure Identification by Mass Spectral Library Searching

Stephen E. Ste
NIST Mass Spectrom

A library-search
its electron-ioniz
retrieves library
compound. It the
structures of the
is weighted acco
"peaks-in-comm
product function
compound is pr
including "neutr
compound and
searching the N
correlations allow
an unknown com
an optimization
improvements th
sure, a relative-
Spectrom 1995, 6,

Estimating Probabilities of Correct Identification From Results of Mass Spectral Library Searches

Stephen E. Stein

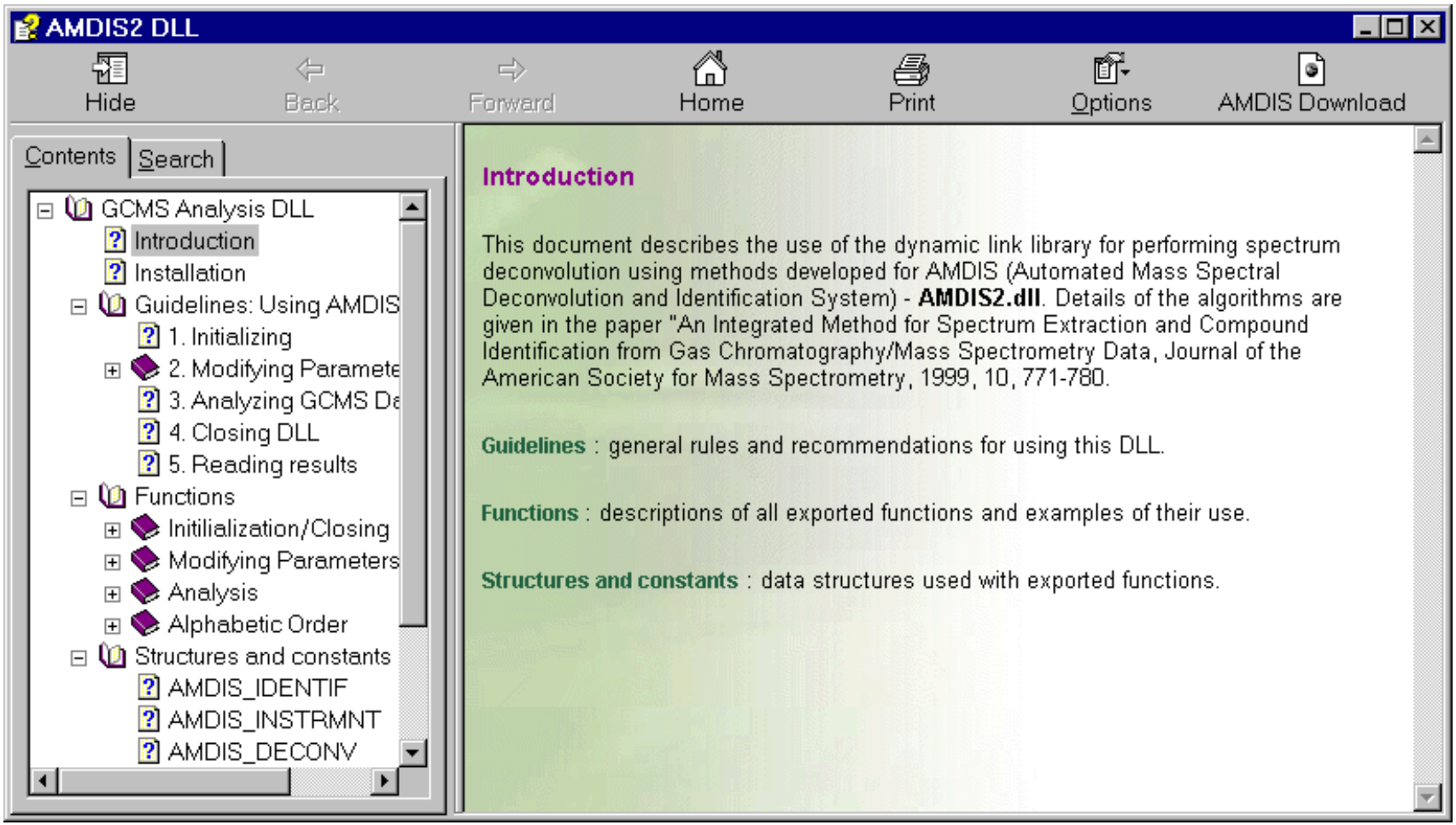
NIST Mass Spectrometry Data Center, National Institute of Standards and Technology, Gaithersburg, Maryland, USA

This work presents a method for using mass spectral match factors reported by library search systems to obtain certain probabilistic indicators of correct identification. The overall probability that a retrieval is correct is formally separated into two independent terms. One of these is the probability that a retrieval is correct assuming that the correct match is contained in the library. This can be computed directly from test results. The other term represents the probability that the spectrum of the unknown compound is actually in the library. While the absolute value of this term cannot be computed, a relative value based solely on search results can be derived. This value may, if desired, be used to refine an initial estimate of the overall probability. Parameters used in this calculation are based on changes in test results caused by the logical removal of the test compounds from the library. These methods were parameterized from results of searching the NIST/EPA/NIH Mass Spectral Database with 12,592 good quality replicate spectra and a simple mass spectral comparison function. The methodology should be equally applicable to other libraries and search systems. *(J Am Soc Mass Spectrom 1994, 5, 316–323)*

Algorithms/Formats

- Search/Analysis Algorithms
  - Available as DLL
- Data Format
  - Provide needed formats
- Data/Algorithm Integration
  - As needed

Packaged Algorithms



Distributors

- Instrument Data Systems (15)

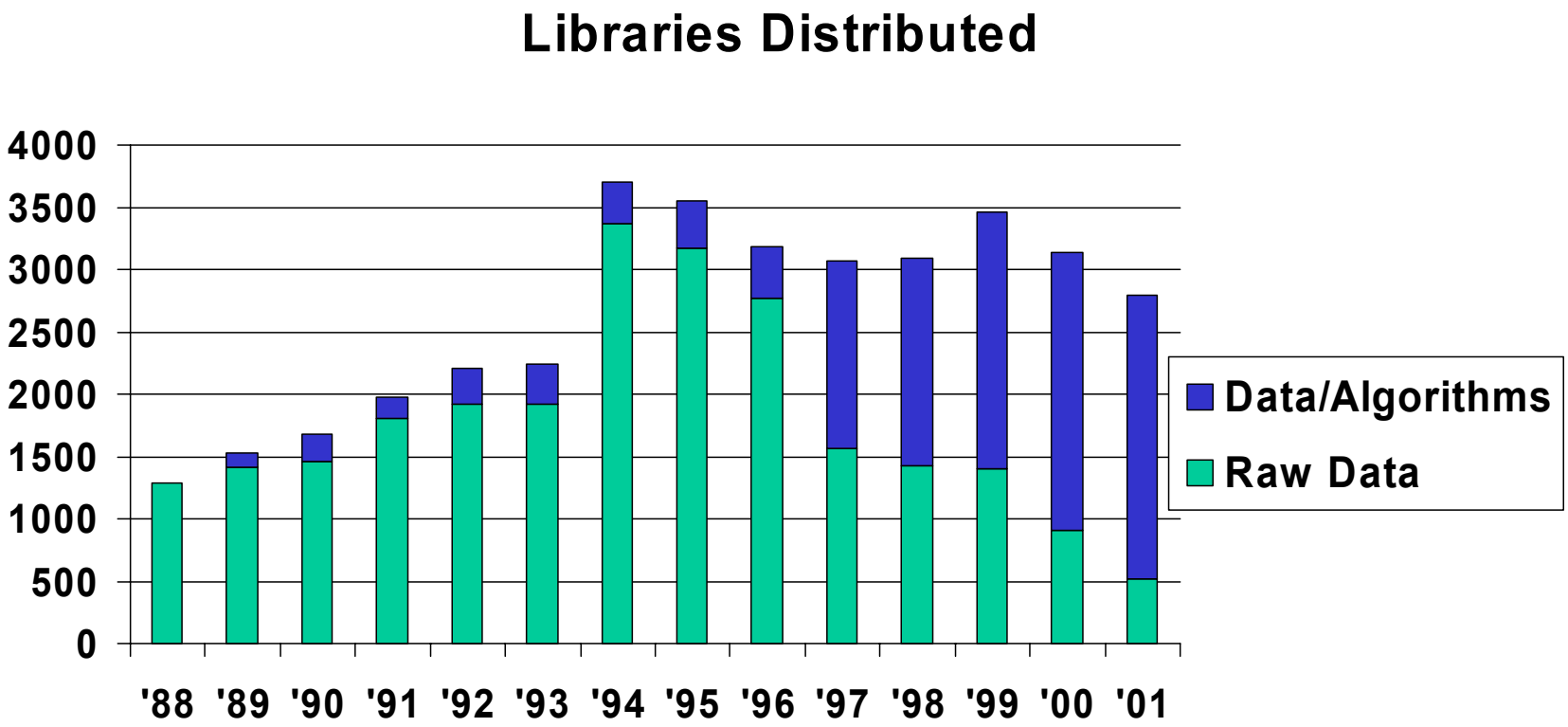
  Agilent, Bruker, ThermoFinnigan, Hitachi, Inficon, JEOL, LECO, Los Gatos, Micromass, MSS, ONIX/Fisons, Perkin-Elmer, Shimadzu, Shrader, Varian/Bear

- Software (21)

  ACD, Aldrich, ARLS, Bio-Rad/Sadtler, Chemical Concepts, ChemSW, ChroMaSoft, CSS, Digital Data Management, Fiveash, Galactic, HD Science, Hiden, JEMS, KORE, Monitor Group, Pro-Lab, Axel Semrau, Spectra Seriea, SIS, Stanton

- Data (3)

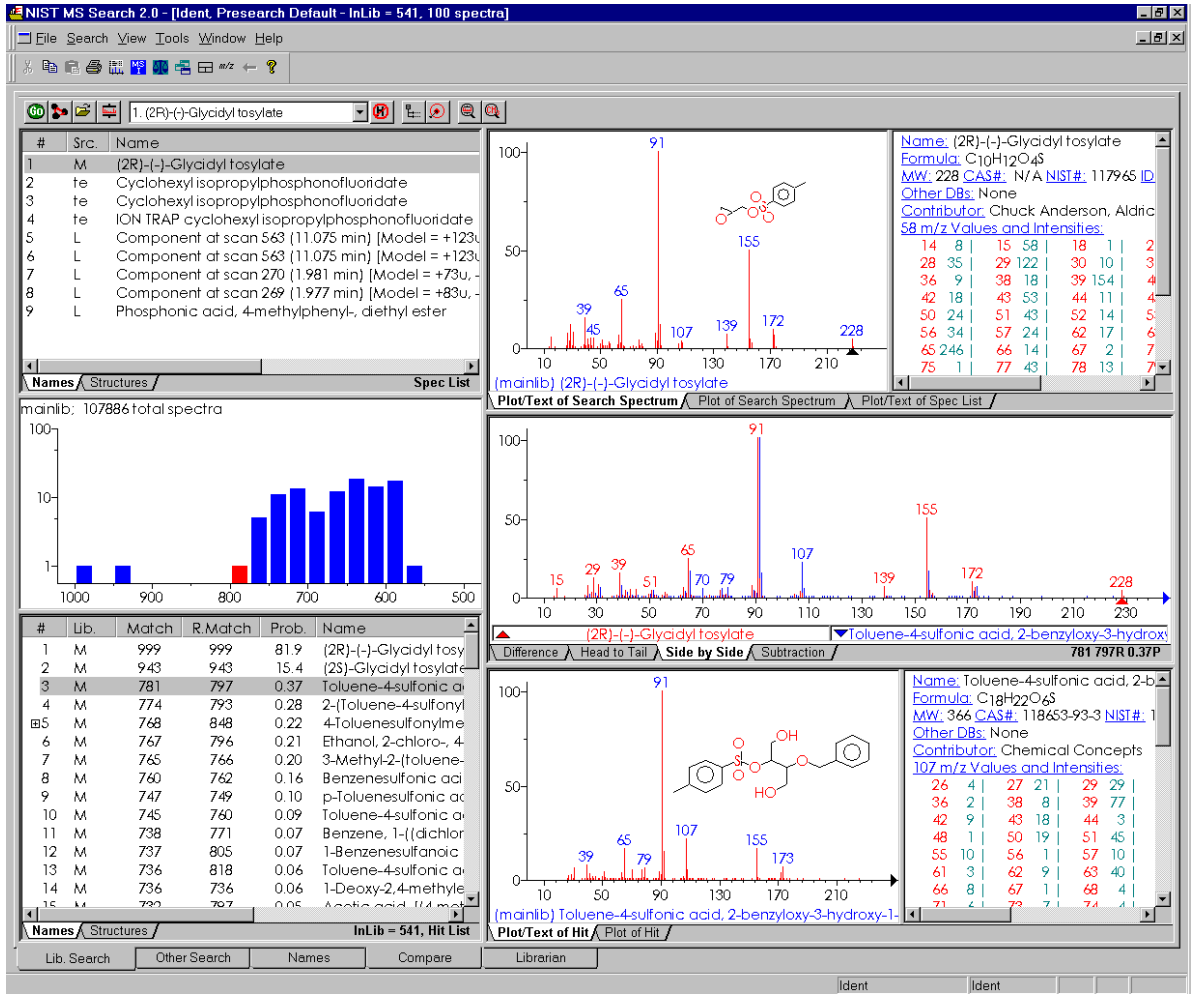
  ERM, JAICI, Wiley


Libraries Distributed
4000
3500
3000
2500
2000
1500
1000
500
0
'88
'89
'90
'91
'92
'93
'94
'95
'96
'97
'98
'99
'00
'01
Data/Algorithms
Raw Data

Field Agent:

Dave Sparkman

NIST Search Program V.2



New Software Policy

- No plans for further GUI development
  - –Freeze Version 2.01 (NIST ’02 release)
  - –Will ‘keep it running’
  - –Distribute freely
- Provide source to distributors
  - –Limited support
  - –Cooperative R&D Agreements possible

- Will Continue
  - –AMDIS
  - –Algorithms (DLL)
  - –Data formats
  - –Assistance in Integration

Pricing/License Issues

Joan Sauerwein

Standard Reference Data Program

NEW PRICING STRUCTURE FOR NIST02