The NIST Mass Spectrometry Data Program

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

Distributors’ Meeting
PITTCION ‘02
March 19, 2002
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,

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.

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.

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.

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).

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


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

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

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?

<table>
<thead>
<tr>
<th>Spectra that produce top matches with NIST 98</th>
<th>within top 10 hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetic acid, 1-methylethyl ester</td>
<td>1-Butanol</td>
</tr>
<tr>
<td>Acetic acid, butyl ester</td>
<td>2-Butenal</td>
</tr>
<tr>
<td>Butanoic acid, 3-oxo-, methyl ester</td>
<td>Ethanol, 2-butoxy-</td>
</tr>
<tr>
<td>1-Butanol</td>
<td>Hexane</td>
</tr>
<tr>
<td>1,4-Cyclohexanedicarboxylic acid, dimethyl ester</td>
<td>Methane, diethoxy-</td>
</tr>
<tr>
<td>1,4-Dioxane</td>
<td>Pentane</td>
</tr>
<tr>
<td>Ethanol, 2-mercaptoethyl ester</td>
<td>1-Propanol</td>
</tr>
<tr>
<td>Ethanol, 2-phenoxyethyl ester</td>
<td>Propanoic acid, 2-methyl-, 2-methylpropyl ester</td>
</tr>
<tr>
<td>Ethyl Acetate</td>
<td></td>
</tr>
<tr>
<td>Glycerin</td>
<td>incognito in the hit list</td>
</tr>
<tr>
<td>Hexanal, 2-ethyl-</td>
<td>1,4-Cyclohexanedicarboxylic acid, dimethyl ester</td>
</tr>
<tr>
<td>2-Hexanone, 5-methyl-</td>
<td>Ethanol, 2-butoxy-</td>
</tr>
<tr>
<td>Methylene Chloride</td>
<td>Ethyl ether</td>
</tr>
<tr>
<td>N-Methyl-N-(2-hydroxyethyl)carbamic acid, methyl ester</td>
<td>Hexane</td>
</tr>
<tr>
<td>2-Propenoic acid, ethyl ester</td>
<td>Octane</td>
</tr>
<tr>
<td>Propanoic acid, 3-ethoxy-, ethyl ester</td>
<td>1-Propanamine</td>
</tr>
<tr>
<td>Triethylamine</td>
<td></td>
</tr>
<tr>
<td>p-Xylene</td>
<td></td>
</tr>
</tbody>
</table>
Example

3.310 min 2-BUTOXY ETHANOL (20NG).FIN
Thiocyanic acid, 2-(2-butoxyethoxy)ethyl ester

41 45 57
100 50 0
60 71 72
43 45 74
63 86 89
71 86 100
75 89 101
86 100 130
41 45 57
100 50 0
60 71 72
43 45 74
63 86 89
71 86 100
75 89 101
86 100 130

Top match
Ion Trap
NIST 98

Correct match
Ion Trap
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
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 Spectrom 199, 10, 287-299) © 1999 American Society for Mass Spectrometry
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 than 10 peaks/spectrum
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
<table>
<thead>
<tr>
<th>CAS:</th>
<th>E:</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>89781</td>
<td>232390</td>
<td>(--)-MENTHOL</td>
</tr>
<tr>
<td>490993</td>
<td>5285</td>
<td>Cyclohexanol, 5-methyl-2-(1-methylethyl)-, (1.alpha.,2.beta.,5.beta.)-</td>
</tr>
<tr>
<td>491010</td>
<td>38796</td>
<td>Cyclohexanol, 5-methyl-2-(1-methylethyl)-, (1.alpha.,2.alpha.,5.beta.)-</td>
</tr>
<tr>
<td>491021</td>
<td>38506</td>
<td>Cyclohexanol, 5-methyl-2-(1-methylethyl)-, (1.alpha.,2.alpha.,5.alpha.)-</td>
</tr>
<tr>
<td>3623516</td>
<td>108354</td>
<td>(1S, 2S, 5R)-(+) -NEOMENTHOL</td>
</tr>
<tr>
<td>3623516</td>
<td>221579</td>
<td>2-Isopropyl-5-methylcyclohexanol</td>
</tr>
<tr>
<td>3623527</td>
<td>151683</td>
<td>(--)-MENTHOL</td>
</tr>
<tr>
<td>3623538</td>
<td>108362</td>
<td>(1S,2R,5R)-(+) -ISOMENTHOL</td>
</tr>
<tr>
<td>15356602</td>
<td>151874</td>
<td>(+)-MENTHOL</td>
</tr>
<tr>
<td>15356704</td>
<td>290746</td>
<td>DL-Menthol</td>
</tr>
<tr>
<td>20747493</td>
<td>207493</td>
<td>Cyclohexanol, 5-methyl-2-(1-methylethyl)-, [1R-(1.alpha.,2.alpha.,5.beta.)]-</td>
</tr>
<tr>
<td>20752345</td>
<td>20752345</td>
<td>Cyclohexanol, 5-methyl-2-(1-methylethyl)-, [1R-(1.alpha.,2.alpha.,5.alpha.)]-</td>
</tr>
<tr>
<td>23283978</td>
<td>108362</td>
<td>(1S,2R,5R)-(+) -ISOMENTHOL</td>
</tr>
<tr>
<td>152083</td>
<td>152083</td>
<td>(+)-ISOMENTHOL</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benzoic acid, 2-(acetyloxy)</td>
<td>25407</td>
<td>“7”</td>
</tr>
<tr>
<td>Benzoic acid, 2-(acetyloxy)</td>
<td>23813</td>
<td>“7”</td>
</tr>
<tr>
<td>Benzoic acid, 2-(acetyloxy)</td>
<td>79540</td>
<td>“7”</td>
</tr>
<tr>
<td>Benzoic acid, 2-(acetyloxy)</td>
<td>22205</td>
<td>“7”</td>
</tr>
<tr>
<td>Benzoic acid, 2-(acetyloxy)</td>
<td>157126</td>
<td>“7”</td>
</tr>
<tr>
<td>Acetylsalicylic acid</td>
<td>119370</td>
<td>“7”</td>
</tr>
<tr>
<td>Benzoic acid, 2-(acetyloxy)</td>
<td>189933</td>
<td>“7”</td>
</tr>
<tr>
<td>Benzoic acid, 2-(acetyloxy)</td>
<td>126014</td>
<td>“7”</td>
</tr>
<tr>
<td>2-Acetoxybenzoic acid</td>
<td>221215</td>
<td>“Sel”</td>
</tr>
<tr>
<td>o-Acetoxybenzoic acid</td>
<td>227751</td>
<td>“7”</td>
</tr>
<tr>
<td>Acetylsalicylic acid</td>
<td>246070</td>
<td>“7”</td>
</tr>
<tr>
<td>Acetylsalicylic acid</td>
<td>246071</td>
<td>“7”</td>
</tr>
<tr>
<td>Acetylsalicylic acid</td>
<td>246073</td>
<td>“7”</td>
</tr>
<tr>
<td>Acetylsalicylic acid</td>
<td>246079</td>
<td>“8”</td>
</tr>
<tr>
<td>Aspirin</td>
<td>247685</td>
<td>“7”</td>
</tr>
<tr>
<td>Aspirin</td>
<td>250572</td>
<td>“R”</td>
</tr>
<tr>
<td>2-Acetylsalicylic acid</td>
<td>257699</td>
<td>“7”</td>
</tr>
<tr>
<td>2-(Acetyloxy)benzoic acid</td>
<td>287701</td>
<td>“7”</td>
</tr>
<tr>
<td>Acetylsalicylic acid</td>
<td>290983</td>
<td>“R”</td>
</tr>
</tbody>
</table>
Replace Mixture Spectrum with “Pure” Spectrum
Algorithms:
Search and Error Discovery


Stephen E. Stein
NIST Mass Spectrometry Data Center
USA

Donald R. Scott
Atmospheric Research and Exposure Science Division
Triangle Park, North Carolina, USA

Five algorithms proposed match spectra of compounds to reference spectra. Algorithms were optimized by varying parameters. The accuracy of the peak polynomial peak-probability libraries was used to determine the angle between spectra. The Euclidean distance was calculated as the square root of the mass-weighting power. Selection of peaks on the results was the basis for identifying peaks in common. The use of many peaks in common was effective.

Chemical Substructure Identification by Mass Spectral Library Searching

Stephen E. Stein
NIST Mass Spectrometry Data Center
USA

A library-search algorithm retrieves library spectra. The algorithm incorporates a list of questions about the structure of the compound. It is weighted according to the mass of the compound. The peaks in common of the mass spectra of the compound and the known structure are searched for. Using the NIST library of known compounds and their spectra, the chemical structure of the unknown can be determined.

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
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 probability of a retrieval is correct assuming that the correct match is contained in the library. This can be computed directly from the results. Other terms in the calculation include 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 previous results.

This work was supported by the National Institute of Standards and Technology. The methodology should be equally applicable to other libraries and search systems. (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

Introduction


**Guidelines**: general rules and recommendations for using this DLL.

**Functions**: descriptions of all exported functions and examples of their use.

**Structures and constants**: data structures used with exported functions.
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
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

NIST98
- Data Only (including structures) $600 per copy
- Data & Software $800 per copy
- Non-spectral Data $125 per copy
- Data Upgrade $350 per copy
- Software Only $250 per copy

NIST02
- Total Database $850 per copy
- Non-spectral Data $250 per copy
- Data Upgrade $500 per copy
- Free