Knihovny spekter
Spectral libraries

Pavel Matějka
Pavel.Matejka@vscht.cz
Pavel.Matejka@gmail.com
www.vscht.cz/anl/matejka

Knihovny spekter

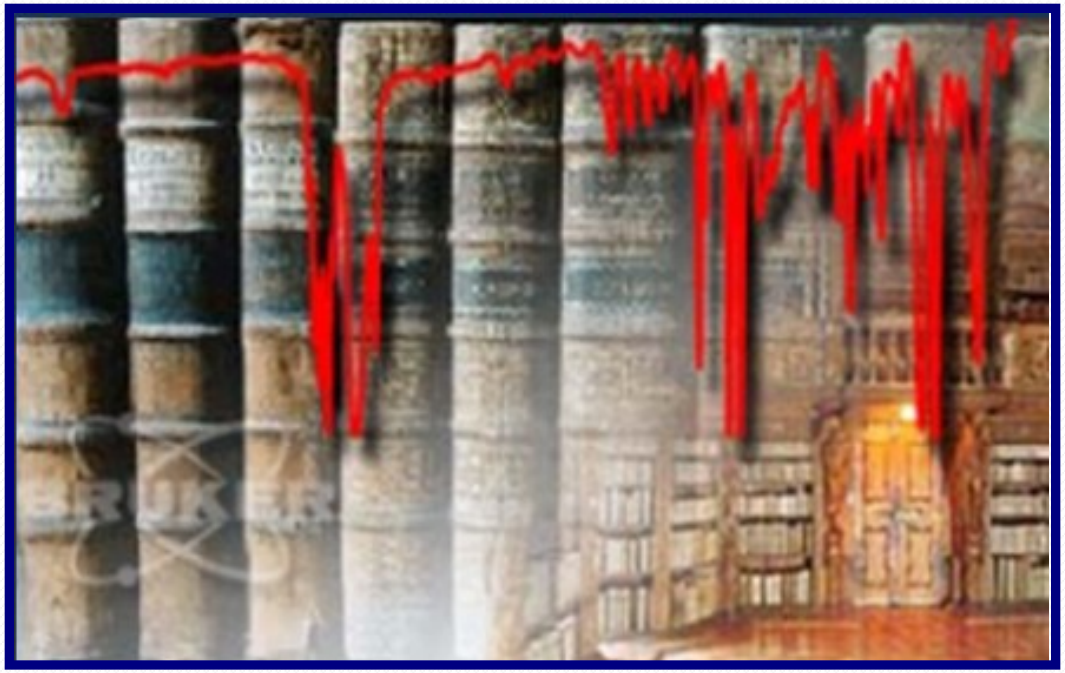
- Tištěné katalogy
  - Schrader, Meyer
- Elektronické knihovny – IR, Raman, MS, NMR, UV-vis, ESCA, Auger, difraktogramy
  - SADTLER standard spectra 54 000
  - COBLENTZ SOCIETY infrared spectra 10 000
  - ALDRICH library of FTIR spectra 30 000
  - HUMMEL polymer library 3 000
  - SIGMA libraries

Knihovny spekter

- https://ftirsearch.com/features/lib_list.htm
- http://speclab.cr.usgs.gov/spectral-lib.html
- http://www.peptideatlas.org/speclib/
- http://www.stjapan.de/
- http://www.swgdrug.org/ms.htm
- http://www.sisweb.com/software/ms/nist.htm
- http://www.nist.gov/srd/nist1a.cfm
- http://peptide.nist.gov/
- http://www.lib.utexas.edu/chem/info/spectra.html
- http://riodb.ibase.aist.go.jp/ssnmr/index_eng.html

Dostupnost spekter – knihovny spekter

- **BRUKER MERCK FT-Raman Library**
  - 3 661 spekter organických látek (3500 – 150 cm^{-1})
  - 353 spekter anorganických látek (3500 – 50 cm^{-1})



Dostupnost spekter – knihovny spekter

- **S.T. Japan Raman Databases for ACD/Labs**
- **Raman ICHEM Database for ACD/Labs**
- Databáze Ramanových spekter – cca 8000 spekter
  - měřeno v „National Institute of Advanced Industrial Science and Technology “ či získáno z knihovny S.T. Japan ICHEM.
  - celá databáze a aplikačně orientované subsety komerčně dostupné v ACD/Labs formátu
  - zahrnuje též polymery a komerční vzorky.
- http://www.acdlabs.com/products/spec_lab/exp_spectra/spec_libraries/st_japan_raman.html
- http://www.stjapan-europe.de/Products/SpectralDatabases/tabid/68/Default.aspx#ram

Tvorba knihoven spekter

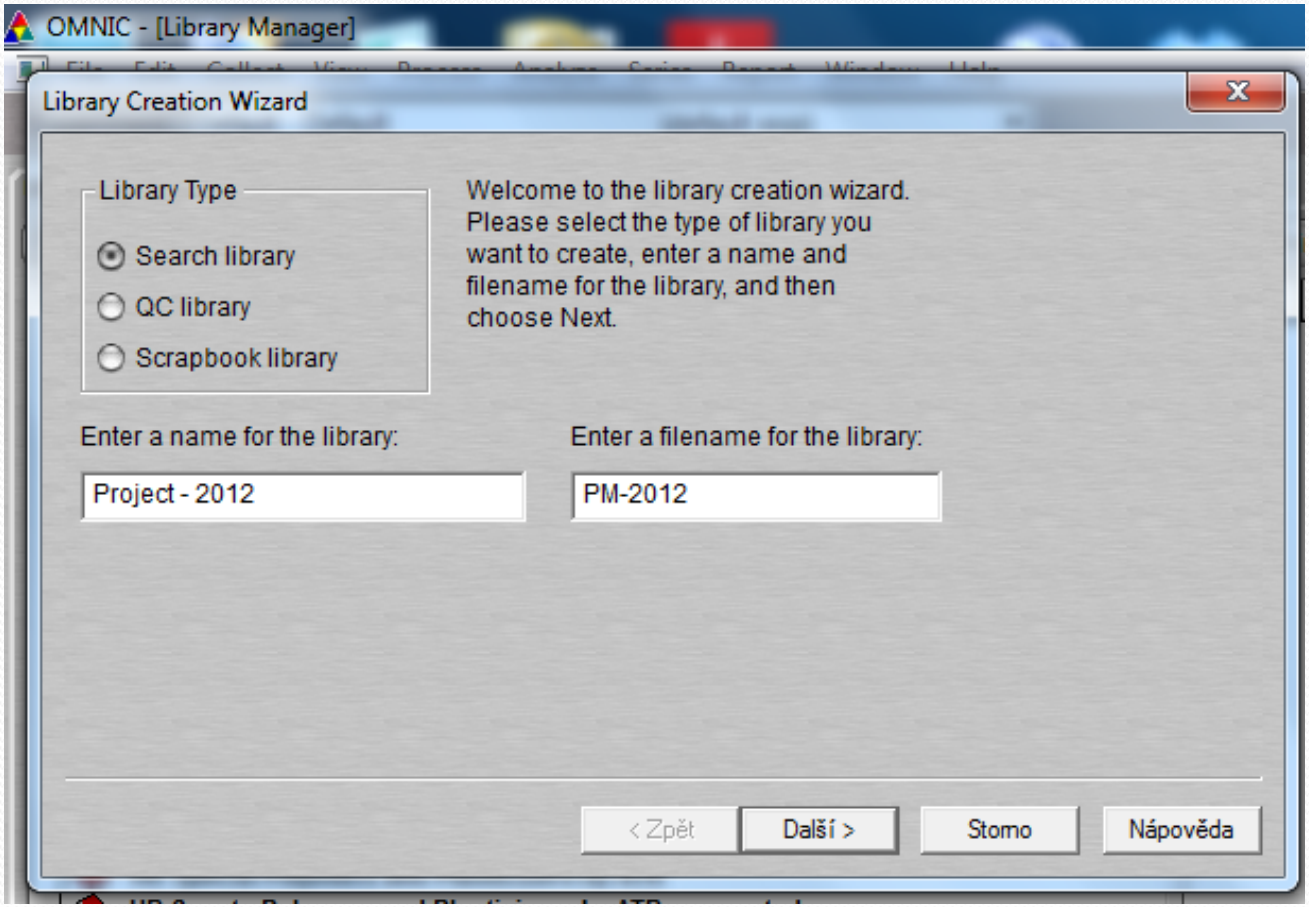
- **Uživatelské knihovny – tvorba z vlastních naměřených spekter**

Úprava spektra

- **Potlačení šumu ?**
- **Srovnání základní linie ?**
- **Stejné (konstantní) rozlišení**
- **Eliminace nežádoucích (nepravých) pásů**
- **Normalizace spektra**

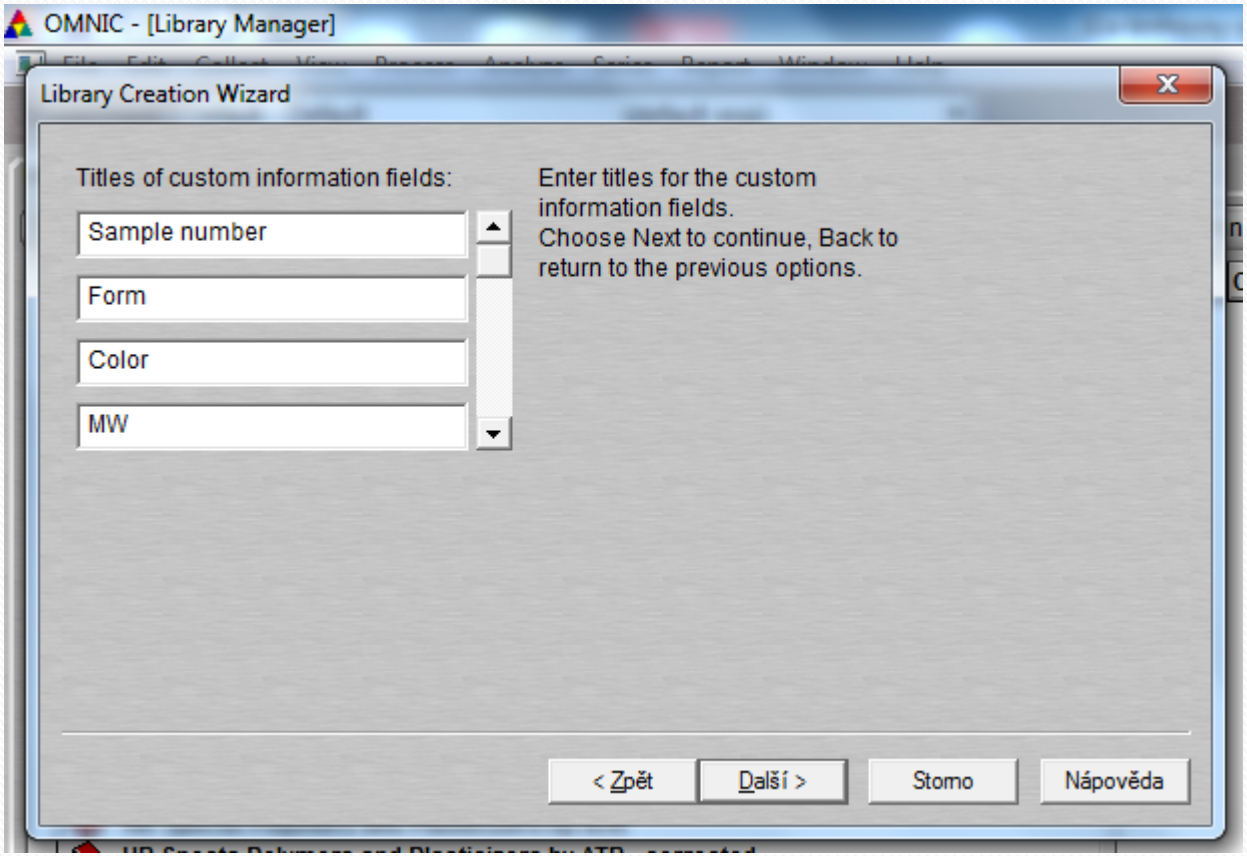
Tvorba knihoven spekter

- **Uživatelské knihovny – tvorba z vlastních naměřených spekter – typ knihovny**



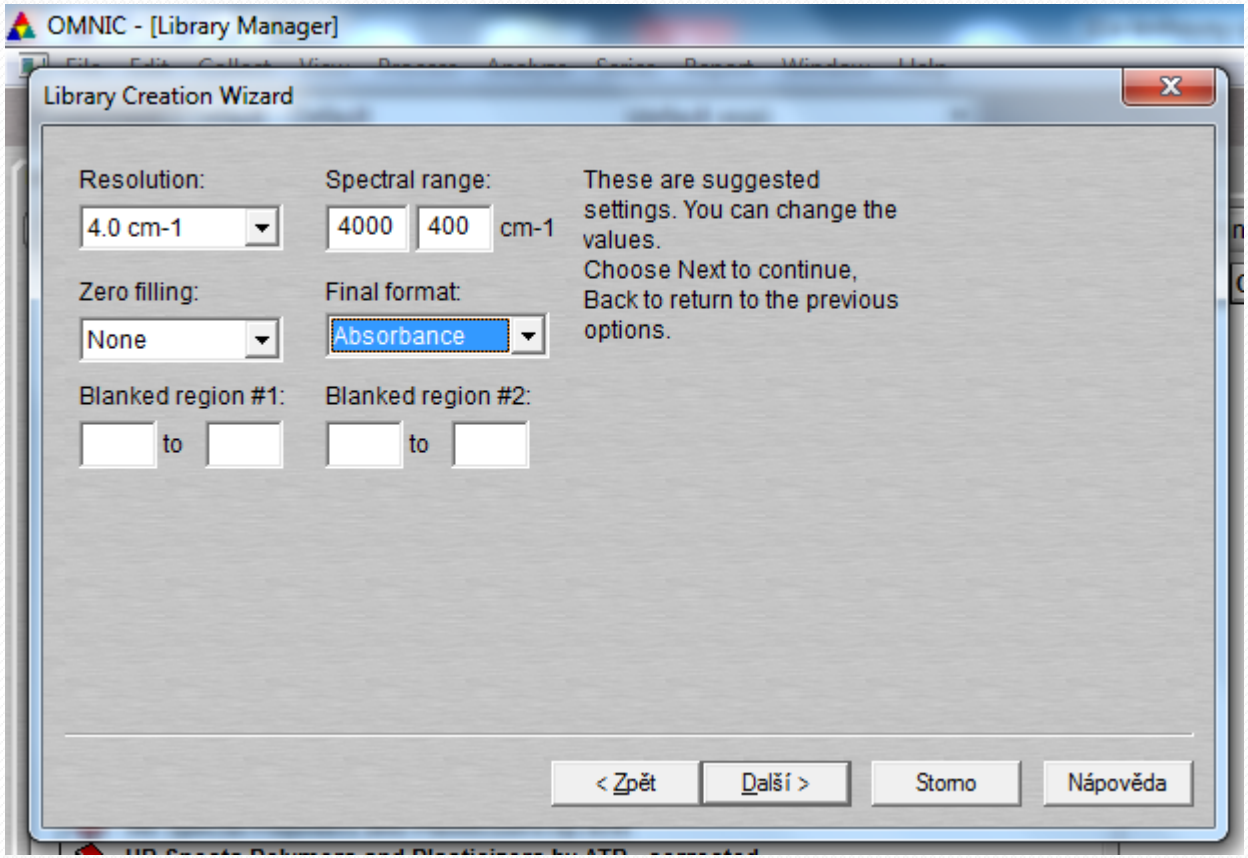
Tvorba knihoven spekter

- **Uživatelské knihovny – tvorba z vlastních naměřených spekter – informační položky**



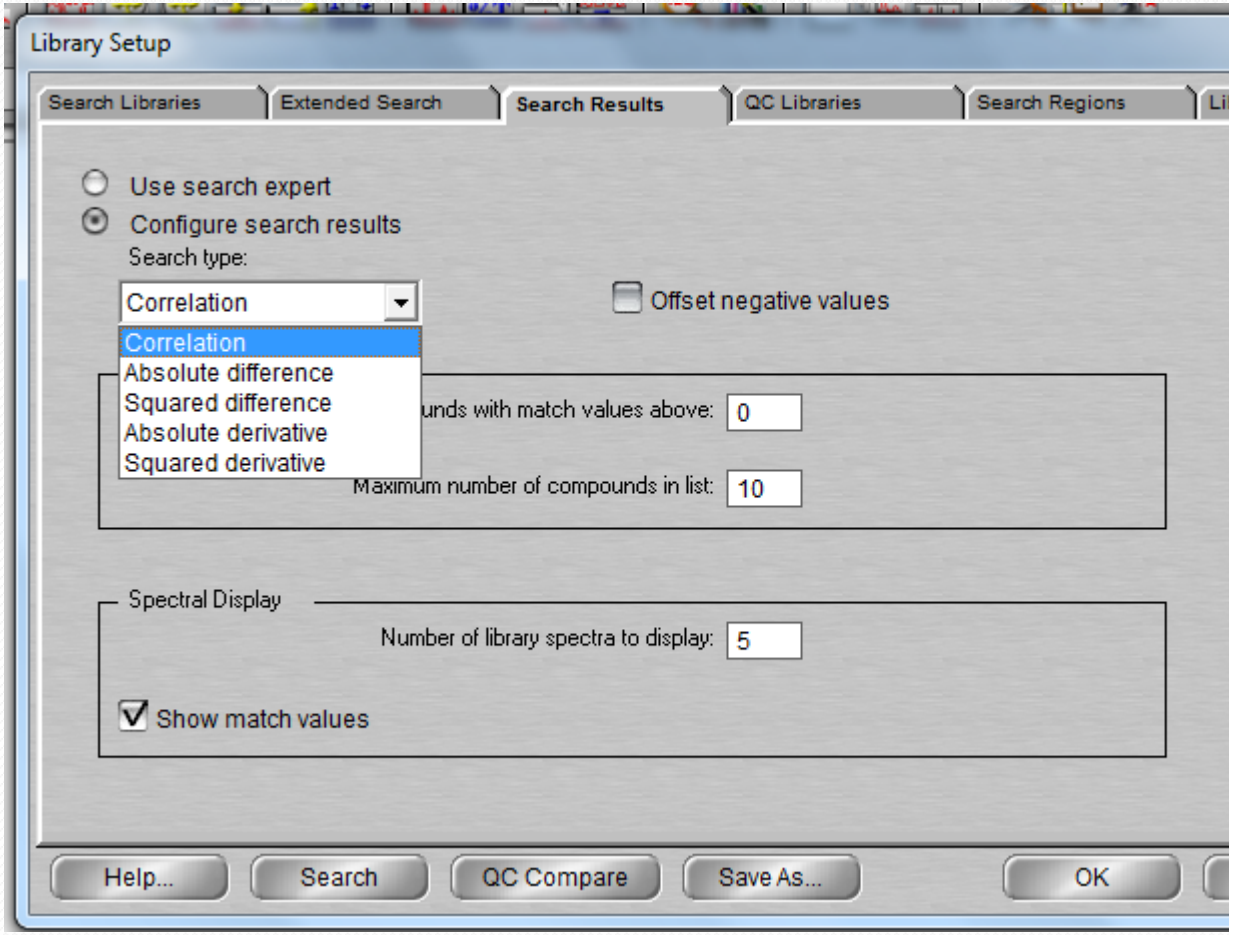
Tvorba knihoven spekter

- **Uživatelské knihovny – tvorba z vlastních naměřených spekter – spektrální parametry**



Vyhledávací algoritmy

- „Black box“ – „expert mode“
- Korelační



$$Y_{A\lambda 1} - Y_{B\lambda 1}$$

$$\left| Y_{A\lambda 1} - Y_{B\lambda 1} \right|$$

$$\left| Y_{A\lambda 1} - Y_{B\lambda 1} \right|^2$$

$$\left| Y'_{A\lambda 1} - Y'_{B\lambda 1} \right|$$

$$\left| Y'_{A\lambda 1} - Y'_{B\lambda 1} \right|^2$$

Vyhledávací algoritmy

Correlation	Normally gives the best results and is recommended for most applications. The algorithm removes any effect of offset in the unknown spectrum, thus eliminating the effects of baseline variation.
Absolute Difference	Puts more weight on the small differences between the unknown spectrum and library spectra. This means that impurities will have a larger effect on the search results.

Vyhledávací algoritmy

Squared Difference	Emphasizes the large peaks in the unknown spectrum. Use this algorithm if you are identifying a noisy spectrum.

Vyhledávací algoritmy

Absolute Derivative	Gives small peaks and **peak shifts** an increased effect on the search results. The algorithm removes any differences between the unknown and library spectra caused by an offset in the unknown spectrum. This algorithm is useful when you want to emphasize **peak positions** rather than peak intensities. Use this algorithm if you are identifying a spectrum with a tilted baseline (and you don't want to correct the baseline before the search).

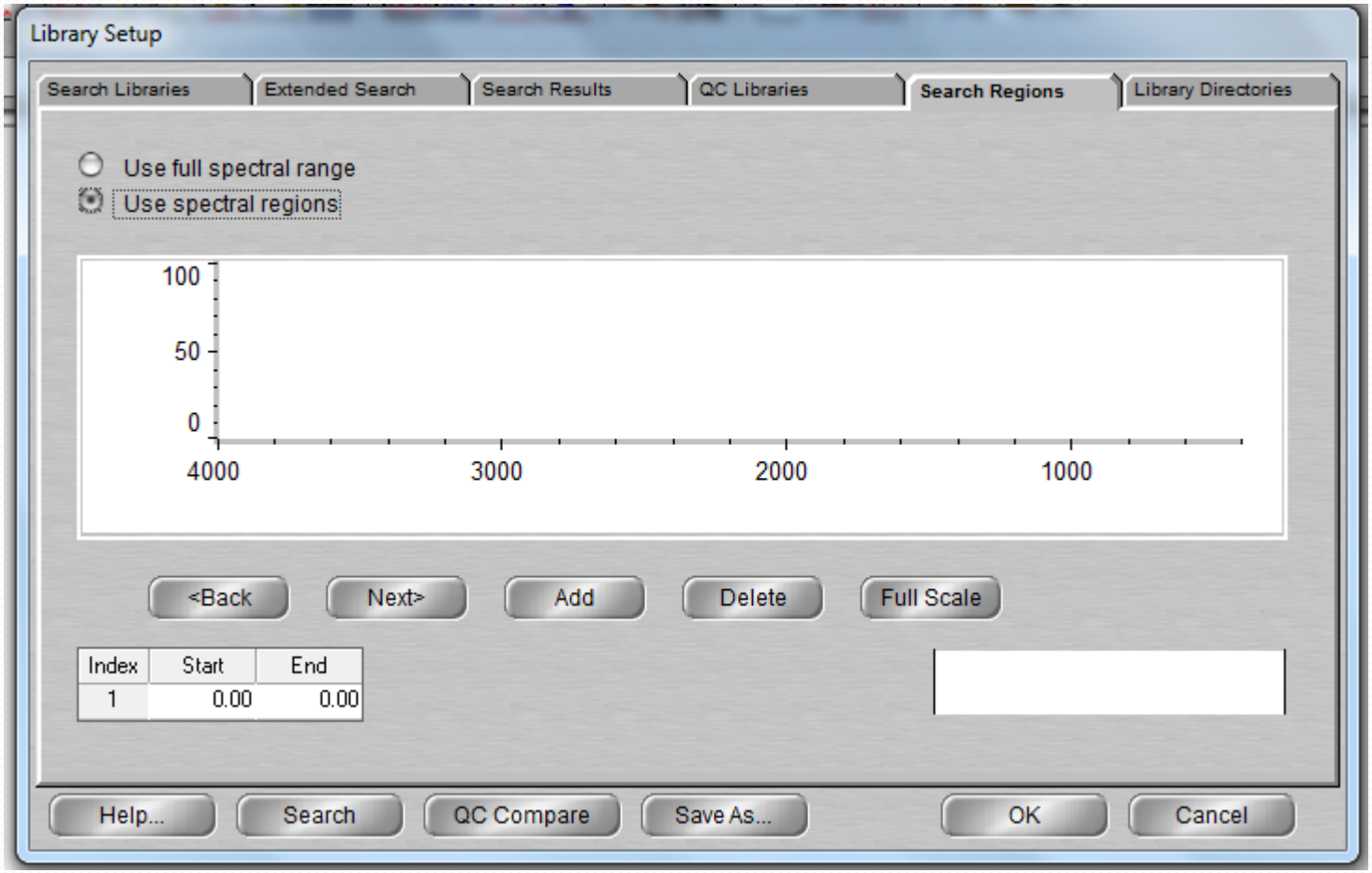
Vyhledávací algoritmy

Squared Derivative	Emphasizes **large peaks** as well as **peak shape**. The algorithm removes any differences between the unknown and library spectra caused by an offset in the unknown spectrum. This algorithm works well with spectra of poor quality.

Další algoritmy – Eukleidovské vzdálenosti
https://ftirsearch.com/help/algo.htm

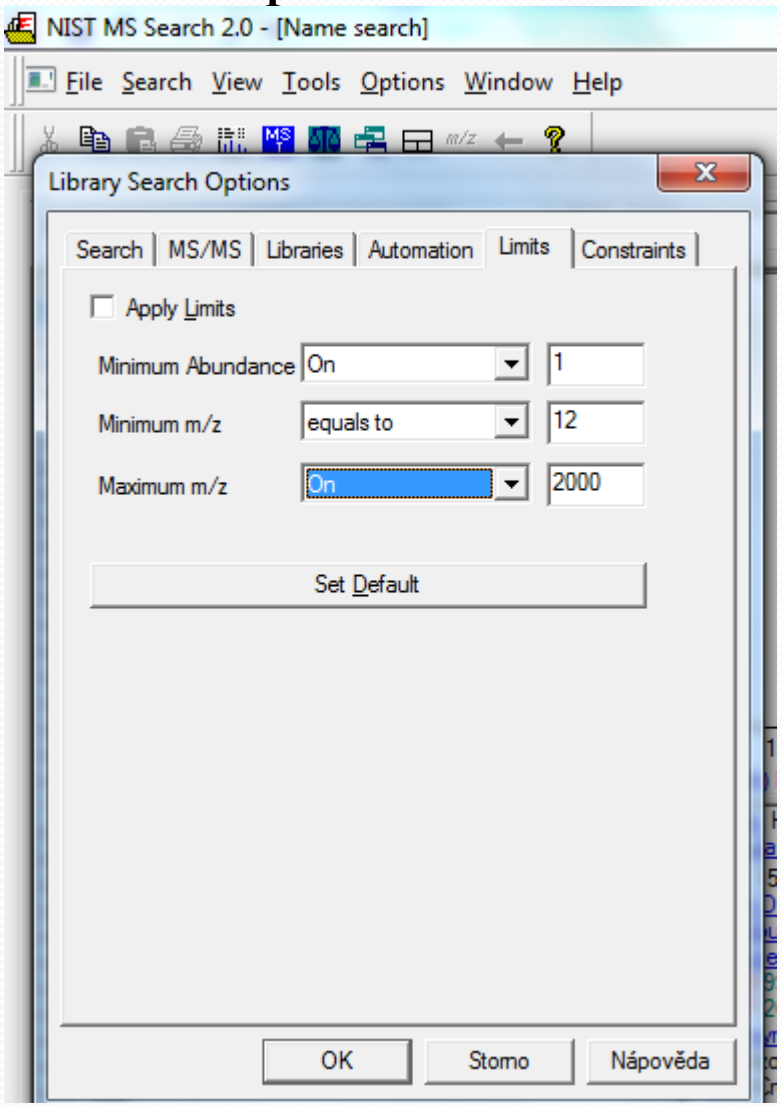
Vyhledávací algoritmy

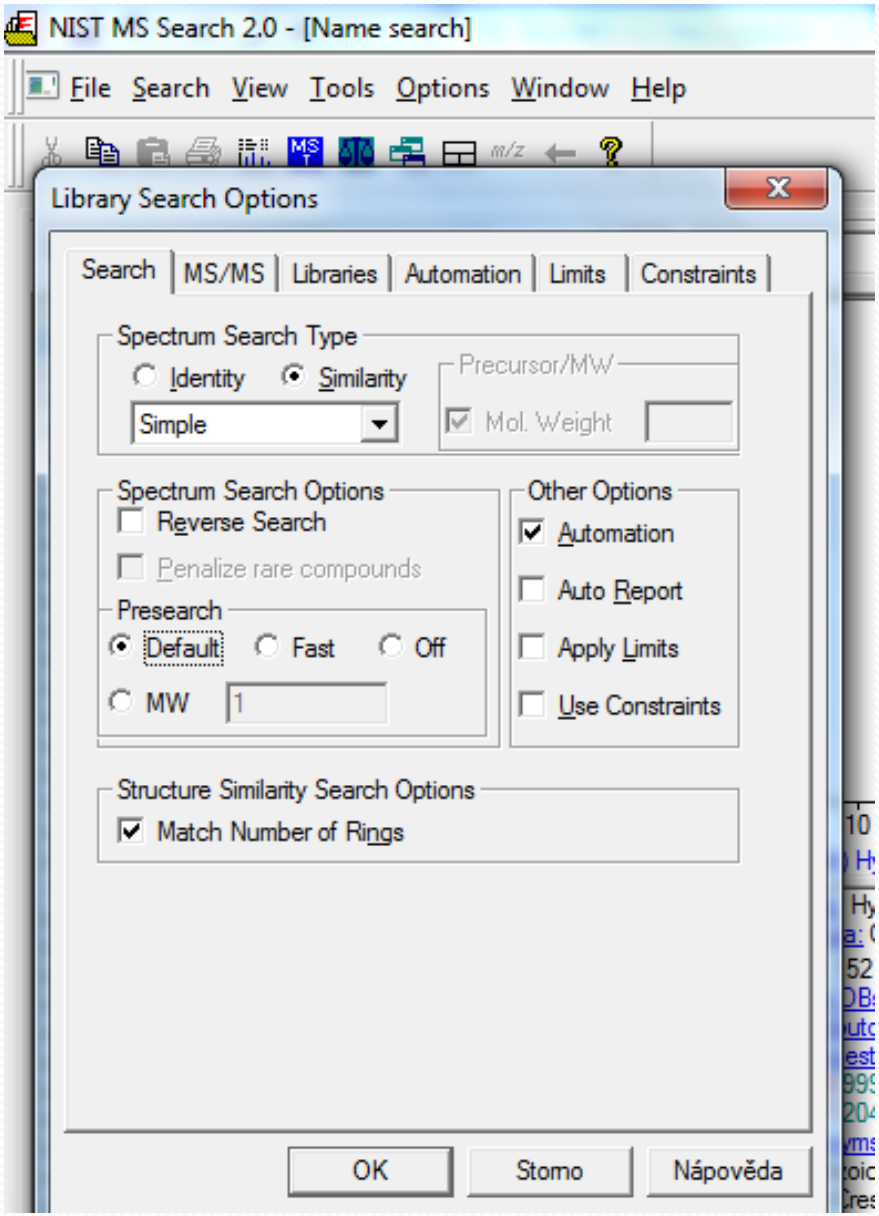
- Volba spektrální oblasti



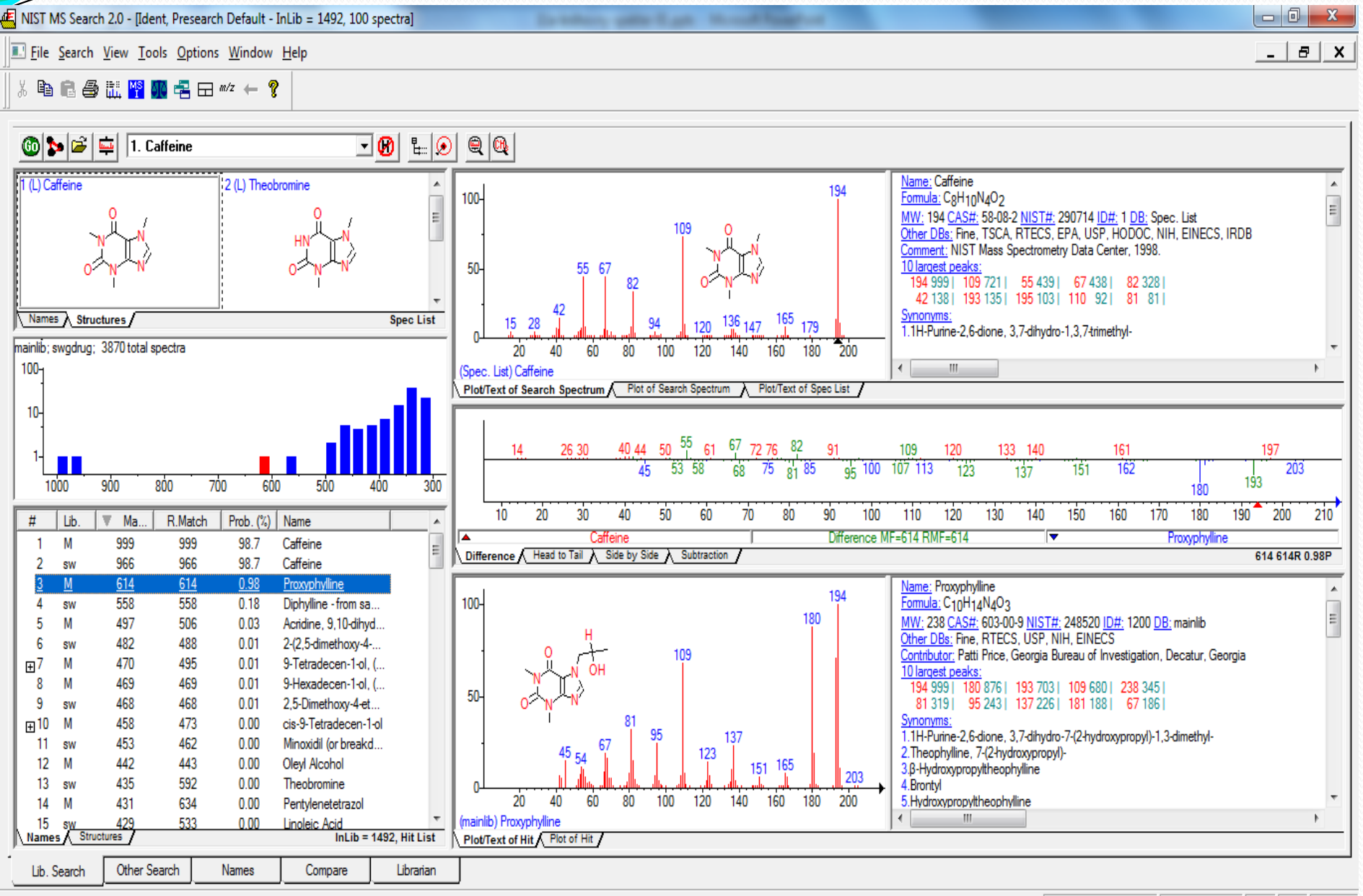
Vyhledávací algoritmy

- Volba parametrů - MS





MS - výsledky


NIST MS Search 2.0 - [Ident, Presearch Default - InLib = 1492, 100 spectra]
File Search View Tools Options Window Help
1. Caffeine
1 (L) Caffeine
2 (L) Theobromine
Names Structures Spec List
mainlib; swgdrug; 3870 total spectra
Lib. Ma... R.Match Prob. (%) Name
1 M 999 999 98.7 Caffeine
2 sw 966 966 98.7 Caffeine
3 M 614 614 0.98 Proxyphylline
4 sw 558 558 0.18 Diphylline - from sa...
5 M 497 506 0.03 Acridine, 9,10-dihyd...
6 sw 482 488 0.01 2-(2,5-dimethoxy-4-...
7 M 470 495 0.01 9-Tetradecen-1-ol, (...
8 M 469 469 0.01 9-Hexadecen-1-ol, (...
9 sw 468 468 0.01 2,5-Dimethoxy-4-et...
10 M 458 473 0.00 cis-9-Tetradecen-1-ol
11 sw 453 462 0.00 Minoxidil (or breakd...
12 M 442 443 0.00 Oleyl Alcohol
13 sw 435 592 0.00 Theobromine
14 M 431 634 0.00 Pentylenetetrazol
15 sw 429 533 0.00 Linoleic Acid
Names Structures InLib = 1492, Hit List
(Spec. List) Caffeine
Plot/Text of Search Spectrum Plot of Search Spectrum Plot/Text of Spec List
Name: Caffeine
Formula: C8H10N4O2
MW: 194 CAS#: 58-08-2 NIST#: 290714 ID#: 1 DB: Spec. List
Other DBs: Fine, TSCA, RTECS, EPA, USP, HODOC, NIH, EINECS, IRDB
Comment: NIST Mass Spectrometry Data Center, 1998.
10 largest peaks:
194 999 | 109 721 | 55 439 | 67 438 | 82 328 |
42 138 | 193 135 | 195 103 | 110 92 | 81 81 |
Synonyms:
1.1H-Purine-2,6-dione, 3,7-dihydro-1,3,7-trimethyl-
Caffeine
Difference MF=614 RMF=614
Proxyphylline
Difference Head to Tail Side by Side Subtraction
614 614R 0.98P
(mainlib) Proxyphylline
Plot/Text of Hit Plot of Hit
Name: Proxyphylline
Formula: C10H14N4O3
MW: 238 CAS#: 603-00-9 NIST#: 248520 ID#: 1200 DB: mainlib
Other DBs: Fine, RTECS, USP, NIH, EINECS
Contributor: Patti Price, Georgia Bureau of Investigation, Decatur, Georgia
10 largest peaks:
194 999 | 180 876 | 193 703 | 109 680 | 238 345 |
81 319 | 95 243 | 137 226 | 181 188 | 67 186 |
Synonyms:
1.1H-Purine-2,6-dione, 3,7-dihydro-7-(2-hydroxypropyl)-1,3-dimethyl-
2.Theophylline, 7-(2-hydroxypropyl)-
3.β-Hydroxypropyltheophylline
4.Brontyl
5.Hydroxypropyltheophylline
Lib. Search Other Search Names Compare Librarian

MS - výsledky

Proteomics 2007, *7*, 655–667 DOI 10.1002/pmic.200600625

RESEARCH ARTICLE

Development and validation of a spectral library searching method for peptide identification from MS/MS

Henry Lam[1], Eric W. Deutsch[1], James S. Eddes[1], Jimmy K. Eng[1, 2], Nichole King[1], Stephen E. Stein[3] and Ruedi Aebersold[1, 4]

$$\hat{I}_j = \frac{I_j}{\sqrt{\sum_j I_j^2}}$$

$$D = \sum_j \hat{I}_{\text{library},j} \hat{I}_{\text{query},j}$$

$$\Delta D = \frac{D_1 - D_2}{D_1}$$

Příklady publikací

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

Stephen E. Stein

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

Donald R. Scott

Atmospheric Research and Exposure Assessment Laboratory, U. S. Environmental Protection Agency, Research Triangle Park, North Carolina, USA

Příklady publikací

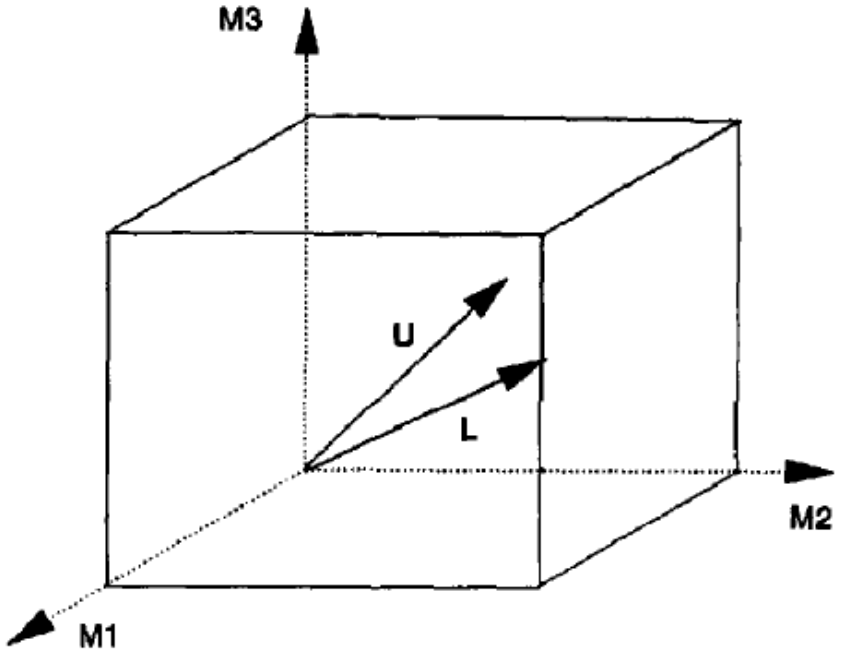


Figure 1. Vector representation of a hypothetical three-peak unknown (U) and library (L) mass spectrum in three-dimensional space (peaks have mass M1, M2, and M3).

Table 1. *Search algorithms investigated*[a]

Euclidean distance

$$\left(1 + \frac{\sum (W_L - W_U)^2}{\sum W_U^2}\right)^{-1}$$

Absolute Value Distance

$$\left(1 + \frac{\sum |W_L - W_U|}{\sum W_U}\right)^{-1}$$

Hertz et al. [9]

$$\frac{\text{average of weighted peak intensity ratios}}{[1 + \text{fraction of unmatched intensities}]}$$

Dot-product (cosine), F_d

$$\frac{\left(\sum W_L W_U\right)^2}{\sum W_L^2 \sum W_U^2}$$

Probability-based matching (PBM) [5b, 5d, 10c]
Uses probability that, by chance, peaks match within an abundance window (W value) by using uniqueness values for mass (U value) and abundance (A) along with a variety of rules and correlation tables.

Composite:

$$\frac{N_U F_D + N_{L\&U} F_R}{N_U + N_{L\&U}}$$

F_D = Dot-Product Term Above
F_R = Ratio of Peak Pairs (below)
W = [Peak Intensity]^n [Mass]^m = Weighted Intensity
N = Number of peaks

$$F_R = \frac{1}{N_{L\&U}} \sum_{i}^{L\&U} \left(\frac{W_{L,i}}{W_{L,i-1}} \frac{W_{U,i-1}}{W_{U,i}}\right)^n$$

where $n = 1$ or -1 when the term in parentheses is less than or greater than unity, respectively

[a] Subscripts: L = library, U = unknown, L & U = peak in both library and unknown spectrum.