Deeper and higher confident annotation of complex metabolomics data by complementary large-scale spectral libraries

Mohamed Elsadig[1], Nikolas Kessler[1], Matthias Szesny[1], Aiko Barsch[1], Ulrike Schweiger-Hufnagel[1]

[1]Bruker Daltonics GmbH & Co. KG
Bremen, Germany

Introduction

- **How orthogonal is the content of commonly used large spectral libraries?**

- **Challenges**

- Even when the same analyte is contained in complementary libraries, their MS/MS spectra might still differ due to different experimental settings, e.g., collision energies.

- Libraries do not provide the same type of meta information (e.g. CAS numbers) which makes comparison of libraries difficult.

- **Solutions**

- Here, MS/MS similarity was used to compare the Bruker NIST 2020 MS/MS library and MetaboBASE 3.0 Personal library (Fig. 1) spectral libraries.

- Furthermore, the chemical space covered by the libraries was assessed - here by determining the chemical similarity. For this purpose, the Tanimoto coefficient between library compounds was calculated. Using this approach, a network was generated with chemical classes clustered (Fig. 2).

Methods

- LC: Elute UHPLC, Intensity Solo C18 column (Bruker).
- MS: timsTOF Pro (Bruker)
- Acquisition: PASEF positive mode
- Software: MetaboScape 2021b (Bruker). Custom data processing was performed using Python, RDKit and Cytoscape.
- Libraries:
  - Bruker NIST 2020 MS/MS library
  - MetaboBASE 3.0 Personal library
- Samples: naphtacene standards obtained from Sigma Aldrich, Germany.

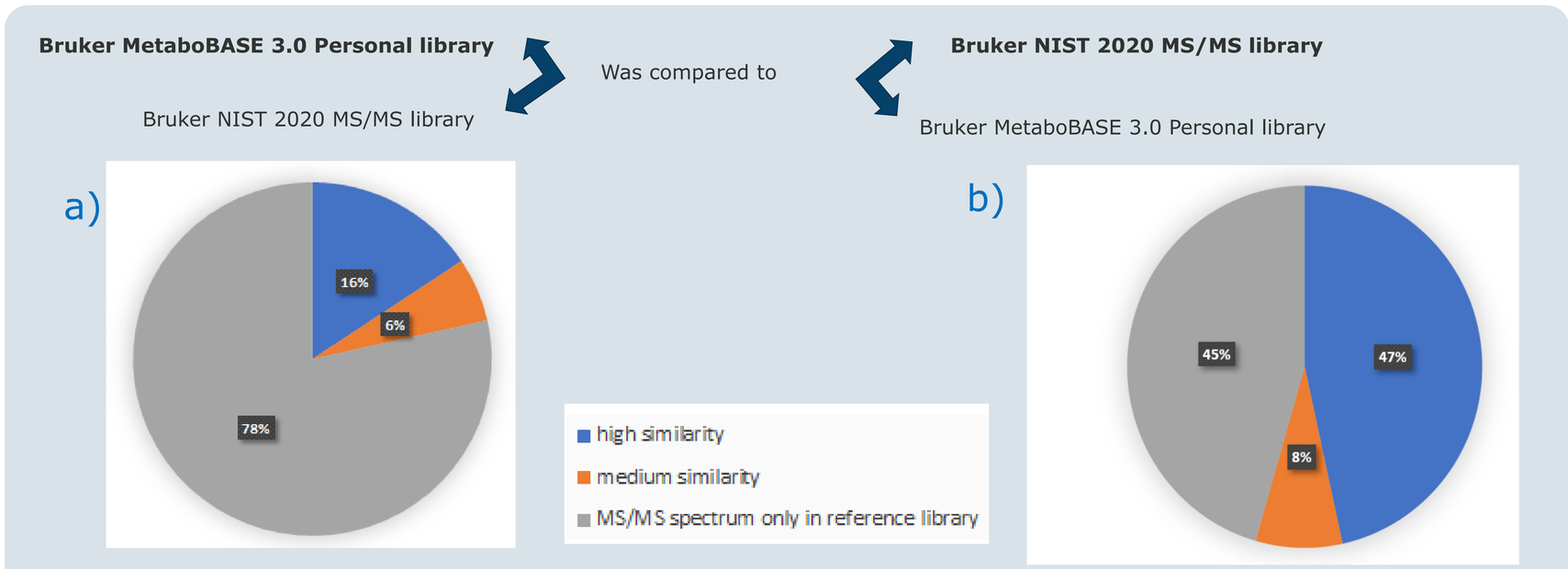


Fig. 1: Spectral matching of compounds between libraries.
Spectra from library 1 was matched against all spectra of library 2 if the precursor molecular formula was identical.
Cosine similarity was used to assess spectral similarity. If multiple scores were obtained, the best match was chosen. Cosine score of >0.8 was considered a high similarity, a score between 0.4 and 0.8 was considered medium, and lower scores were not considered.
a) represents all compounds from MetaboBASE 3.0 (1 x 10^5 compounds)
b) represents all compounds from NIST 2020 (2.8 x 10^4 compounds).

The two libraries contain many unique compounds (78% and 45%, respectively). This indicates that the two libraries, when applied to real world samples, would result in mostly orthogonal and higher total number of annotations. This hypothesis was confirmed for urine samples, where the number of annotated features increased >120 % when using both instead of only on spectral library.

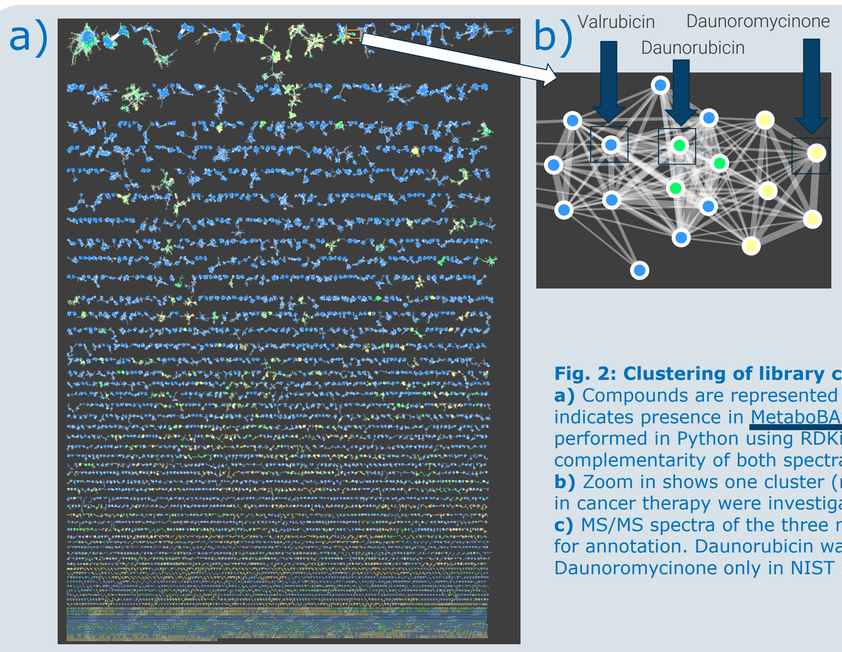
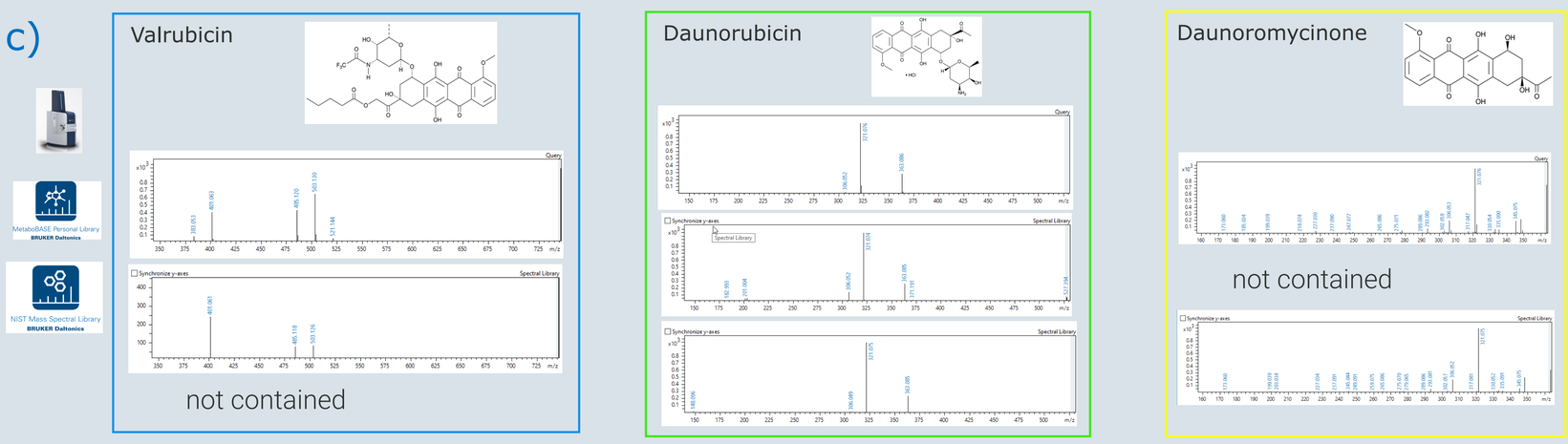


Fig. 2: Clustering of library compounds based on their chemical similarity (Tanimoto coefficient).
a) Compounds are represented as nodes (circles), structurally similar compounds are clustered. Color coding indicates presence in MetaboBASE 3.0 (blue), NIST 2020 (yellow) or in both libraries (green). Calculation was performed in Python using RDKit[2] with a similarity cutoff of 0.7. This visualization highlights the complementarity of both spectral libraries and the large covered chemical space.
b) Zoom in shows one cluster (naphtacenes) with high structural similarity. Three compounds with importance in cancer therapy were investigated in more detail.
c) MS/MS spectra of the three naphtacene compounds were acquired and submitted to spectral library search for annotation. Daunorubicin was found in both spectral libraries, Valrubicin only in MetaboBASE 3.0 and Daunoromycinone only in NIST 2020.



10x Improved spectral library search speed in MetaboScape 2021b

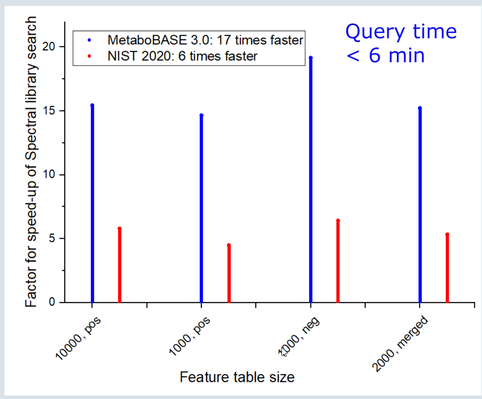


Fig.3: Comparison of Spectral library search speed.
The massive increase in size of spectral libraries requires highly performant search algorithms. The search speed of MetaboScape 2021b was compared to former versions. Feature tables differing in size and polarity were investigated.

In average, a 10 times increased was observed.

Summary

- MetaboBASE 3.0 and NIST 2020 spectral libraries are complementary.
- Both libraries show a good MS/MS match against experimental data.
- MetaboScape enables fast search, even for large spectral libraries.
- This increases annotation coverage in non-targeted metabolomics experiments.

4D-Metabolomics

References

[1] Ralaivola L. et al. (2005) Neural Networks 18(8): 1093-1110

[2] RDKit: Open-source cheminformatics; http://www.rdkit.org

[3] Shannon P. et al. (2003) Genome Res. 13(11): 2498-504

© 2022 Bruker

For Research Use Only. Not for use in diagnostic procedures.

Innovation with Integrity