

CCSPredict: Using a Machine Learning Approach for higher Confidence in Lipid Identification



ASMS 2019, ThP 395

Matthias Szesny¹, Sebastian Wehner¹, Heiko Neuweiger¹, Ulrike Schweiger-Hufnagel¹, Sven W. Meyer¹, Aiko Barsch¹, Nikolas Kessler¹, Lucy A. Woods¹

¹Bruker Daltonik GmbH, Bremen, Germany

Introduction

4D Lipidomics: Trapped ion mobility (TIMS) mass spectrometers offer new options for higher confidence in annotations of target molecules. First, with the additional separation dimension compounds co-eluting from LC columns can be separated. This can result in cleaner MS/MS spectra – crucial for any ID in lipidomics or other small molecule workflows. Moreover, TIMS enables the determination of collisional cross sections (CCS) of ions. These values are specific properties for any ion species under given conditions (type of gas, pressure, temperature). Therefore, acquired CCS values provide increased confidence in compound identification if compared to values from literature or to predicted values generated using recently developed machine learning (ML) algorithms.

Methods

Lipid Standards # 330708 Differential Ion Mobility System Suitability Lipidomix® Kit (from Avanti Polar Lipids, Inc.) were dissolved in Methanol:Dichloromethane (9:1). Milk samples were extracted using a modified Bligh & Dyer [1] method. LC separation was performed using a Bruker Elute UHPLC system (20 min gradient program). MS data were acquired on Bruker timsTOF and timsTOF PRO instruments with and without TIMS separation in ESI positive and negative autoMS/MS modes. For TIMS data acquisition settings of 1M and 5 M Ion Charge Control were applied for timsTOF and timsTOF Pro, respectively. CCS recalibration was performed using TuneMix as calibrant (322 m/z, 152.8; 622 m/z, 201.6; 922 m/z, 241.8; 1221 m/z, 279.9). The resulting data sets were processed in the MetaboScape software (Bruker Daltonics) using the Time aligned Region complete eXtraction (T-ReX™) algorithm. T-ReX 3D was applied to LC-MS/MS data and T-ReX 4D to 4-dimensional LC-TIMS-MS/MS data. Statistical analysis, molecular formula annotation, MS/MS spectral library queries using LipidBlast [2, 3, 4] and prediction of CCS values were conducted in the same integrated client/server software solution. Lipid prediction is based on a support vector regression based machine learning approach described by Zhou et al. [5].

Acknowledgements

We thank Prof. Peter Hoffmann and Dr. Mark Condina (Future Industries Institute, The University of South Australia) for providing the lipid milk extracts.

timsTOF systems generate highly reproducible lipid CCS values

- 10 consecutive measurements of 24 lipid ions on a timsTOF Pro showed highly reproducible CCS values (see Figure 1). The average RSD for all lipids was 0.17%. Minimum and maximum RSD observed were 0.07% and 0.49%, respectively.
- Measurements for the same lipid standard mix on a timsTOF instrument (n=3 runs) demonstrated a high instrument to instrument reproducibility of CCS values (see Figure 2).

4D Lipidomics: Lipid CCS prediction in MetaboScape for increased confidence in lipid identification

- CCS values for known standards were measured on the timsTOF Pro instrument (Figure 1). These were very similar to values predicted by a ML algorithm in MetaboScape (Figure 3). Note that apart from LysoPC 18:1, SM 18:1, Cer 18:1, TG 18:1 the measured lipid species were not used for generating the lipid prediction model.
- CCS values match especially well for those lipid classes which were used for training the lipid prediction model (see Figure 3).

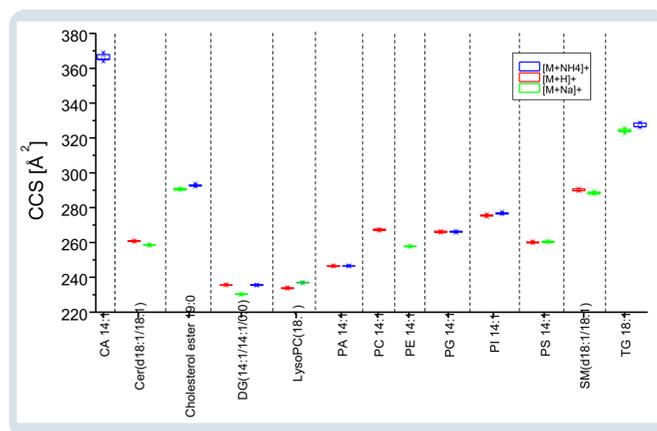


Fig. 1: Highly reproducible CCS values: timsTOF Pro
Box plots of CCS values from 24 ion species (corresponding to 13 lipids standards from Avanti # 330708) measured in 10 consecutive timsTOF Pro runs.

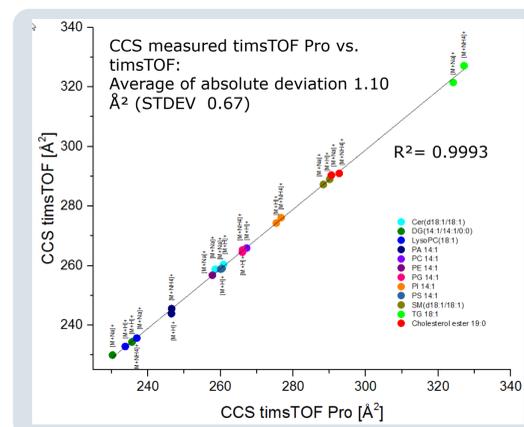


Fig. 2: CCS values measured on timsTOF vs. CCS values measured on timsTOF Pro
CCS values from 23 ion species (corresponding to 12 lipids standards from Avanti # 330708) measured on a timsTOF [average of n=3] are plotted against the corresponding CCS values measured on timsTOF Pro [average of n=10].

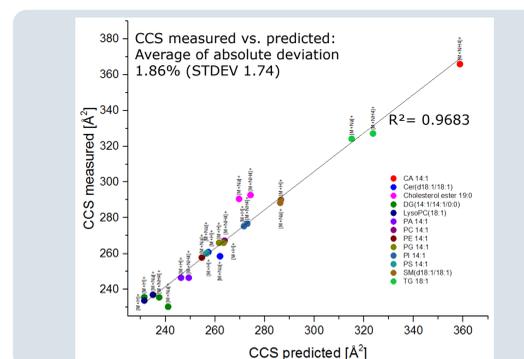


Figure 3: CCS values measured vs. CCS values predicted in MetaboScape 4.0
Average CCS values [n=10] from 24 ion species measured on timsTOF Pro are plotted vs. the CCS values predicted using a machine learning algorithm implemented in MetaboScape 4.0. The lipid class of cholesterol esters (pink) is not covered by the CCS prediction model.

References

- Bligh E. and Dyer W.; Can J Biochem Physiol 1959, 37(8):911-7
- Kind T. et al.; Nature Methods 2015, 10:755-758
- Kind T. et al.; Anal. Chem. 2014, 86(22):11024-11027
- Tsugawa H. et al.; J. of Cheminformatics 2017, 9(19)
- Zhou Z. et al.; Anal. Chem. 2017, 89(17): 9559-9566

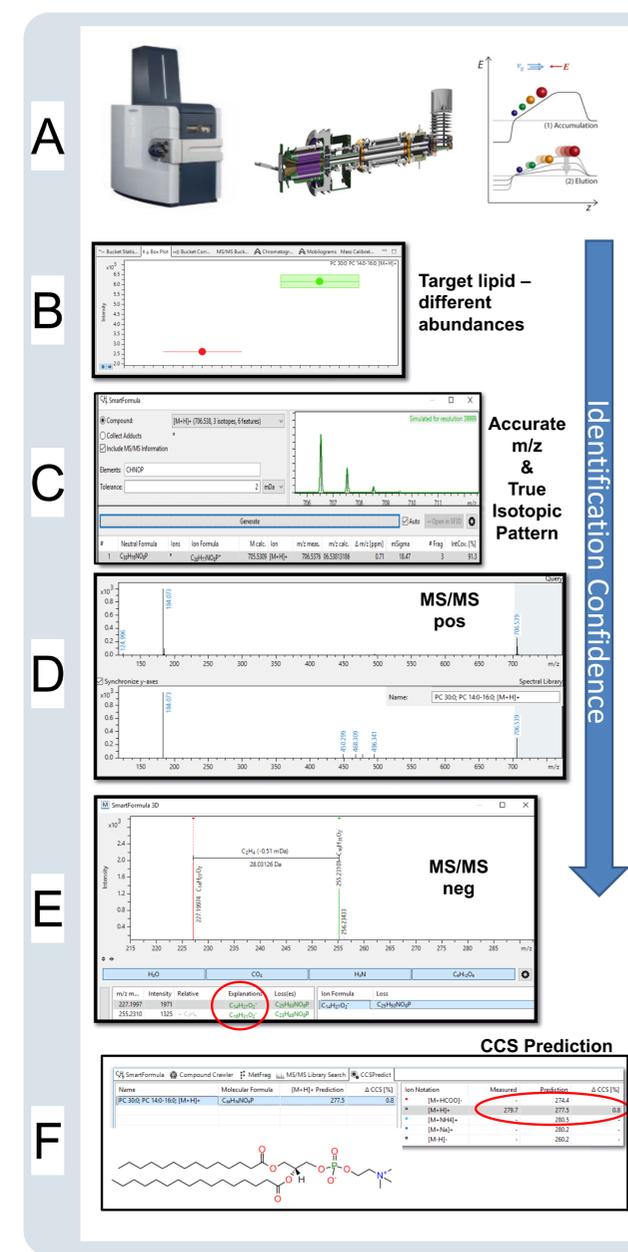


Fig. 4: High confidence lipid identification workflow: A real-world example using MetaboScape 4.0

- Figure 4 highlights the workflow for high confidence lipid identification using MetaboScape 4.0 based on a real-world example.

A) Lipid extracts from two milk samples were analyzed by LC-MS/MS and LC-TIMS-MS/MS using a timsTOF. Shown are the instrument schematic and TIMS operation from ion accumulation to serial elution, where E represents electrical field and vg the gas flow.

B) Box plot of a lipid showing different abundances in two milk lipid extracts detected in ESI positive mode.

C) Assignment of molecular formula based on accurate mass and true isotopic pattern using SmartFormula™ – C₃₈H₇₆N₀O₈P.

D) MS/MS query for ESI positive mode against LipidBlast library returns PC 30:0 as likely candidate. Characteristic head group fragment (184.07 m/z) substantiated assignment as glycerophosphocholine (PC).

E) SmartFormula3DTM algorithm: Fatty acid chains C14:0 and C16:0 could be assigned based on automatic formulae assignment for characteristic fragment ions in negative mode.

F) Very low deviation (0.8 %) between measured vs. predicted CCS values (279.7 vs. 277.5) provided further orthogonal confidence for the identification of the target lipid as: PC 14:0_16:0.

Conclusions

- timsTOF Pro instruments provide highly reproducible CCS values (0.17% RSD).
- CCS values can be determined reproducibly on timsTOF and timsTOF Pro instruments (the two instruments agree on average within 1.1 Å²).
- MetaboScape enables prediction of lipid CCS values based on machine learning.
- CCS values increase confidence in lipid annotation as orthogonal information in addition to accurate mass, true isotopic pattern, MS/MS and retention time information.
- MetaboScape provides a fully integrated solution for confident lipid annotation enabled by 4D Lipidomics.

4D Lipidomics timsTOF Pro