Non-targeted metabolomics ID workflows providing custom annotation confidence reporting: CCS-enabled annotation workflows

ASMS 2021 - FP 357

Aiko Barsch[1]; Matthias Szesny[1]; Ulrike Schweiger-Hufnagel[1]; Sofie Weinkouff[1], Nikolas Kessler[1]

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

Overview

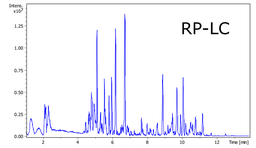
Lack of exhaustive repositories providing retention time, MS/MS spectra and CCS values hamper the annotation and identification (ID) of target compounds in Metabolomics research. Researchers require a solution that automatically and transparently annotates features. Here we present a solution for automatic annotation of targets with up to 5 confidence criteria. Additionally, tentative annotation of knowns and predicted known compounds is supported by automatic *in-silico* fragmentation and CCS prediction based on the novel CCS-Predict Pro model. Paired with customizable annotation quality scoring and visualization the presented annotation workflows enable researchers to assess and report ID level confidence suitable for the study, as recommended by Schymanski *et al.* [1].

Methods

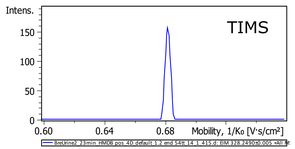
- **Sample:** Human Urine; centrifuged and filtered (0.22µm), 2µl injected, 3 replicates
- **LC:** Elute UHPLC, Intensity Solo C18 column (Bruker).
  - Gradient: Acetonitrile / Water based LC gradient according to T-ReX LC-QTOF Solution (Bruker), allows matching of retention times for >600 compounds
- **MS:** timsTOF Pro 2 (Bruker)
- **Acquisition:** PASEF positive mode
- **Software:** MetaboScape 2022, preliminary Version (Bruker).
- **Libraries:**
  - Bruker HMDB Metabolite Library 2.0
  - Bruker MetaboBASE Personal Library 3.0
  - Bruker NIST 2020 Mass Spectral Library
- **CCS reference values:** CCS Compendium [2]

A) 4D-Metabolomics™: RP-LC-TIMS-MS/MS featuring PASEF provides five indicators of data quality

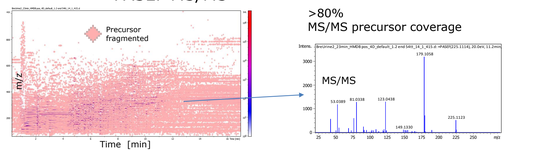
LC: Elute UHPLC
MS: timsTOF Pro 2



1. Retention time
2. Accurate mass
3. True isotopic pattern



4. **TIMS** (Trapped Ion Mobility Separation) = reproducible **CCS values**



5. **PASEF®** data acquisition provides clean MS/MS spectra by **MOMA** = Mobility Offset Mass Aligned

B) High confidence annotation of knowns

TL Enabled by Target List containing CCS values for metabolites derived from CCS Compendium [2], appended by RT and MS/MS

ID criteria	m/z	Isotopic pattern	RT	CCS	MS/MS
Target List	✓	✓	Reference	Reference	Reference

Unified CCS Compendium [2] — 1438 compounds

RTs for >600 compounds

Unified CCS Compendium [2] — CCS values for 1438 compounds

> 800 | > 100.000 | > 27.000 MS/MS spectra for compounds

Name	Molecular Formula	RT [min]	CCS [M+Na]+	CCS [M+H]+	CCS [M-H]-	Library Compound
N-Acetylserotonin	$C_{12}H_{14}N_2O_2$	6.44	153.8	148.8	150.6	N-Acetylserotonin
5-Hydroxyindoleacetate	$C_{10}H_9NO_3$	6.29	148.6	141.1	137.3	5-Hydroxyindoleacetic acid
D-Glucuronolactone	$C_6H_8O_7$	6.13	146.0			D-Glucurono-6,3-lactone
3,4-Dihydroxyphenylacetate	$C_8H_8O_4$	5.73				3,4-Dihydroxybenzeneacetic acid
⋮						
Acetaminophen	$C_8H_9NO_2$	5.41		131.4	132.7	Acetaminophen

Precursor accurate mass matching | Isotopic Pattern matching | RT matching | CCS matching | MS/MS matching

Name	Molecular Formula	AQ	AQ Details	Δm/z [ppm]	mSigma	ΔRT	ΔCCS [%]	MS/MS score
Acetaminophen	$C_8H_9NO_2$			1.394	6.1	0.14	0.0	902.2
Phenylacetyl-L-glutamine	$C_{13}H_{16}N_2O_4$			1.899	9.5	0.16	0.8	969.4
Xanthurenate	$C_{10}H_7NO_4$			1.437	9.2	0.16	0.6	747.0
5'-Methylthioadenosine	$C_{11}H_{15}N_5O_3S$			-0.083	8.8	0.19	1.3	947.9
3-Methyl-L-histidine	$C_7H_{11}N_3O_2$			-0.779	9.0	0.16	1.2	918.9
⋮								

Structure of annotated compound, here Acetaminophen, enables triggering of metabolite prediction for this drug compound (see D).

Annotation Quality Scoring (AQ): Access the confidence in compound annotation with a glimpse!

Higher confidence / Lower confidence

1 Mass accuracy
2 RT deviation
3 Isotopic pattern
4 MS/MS score
5 CCS deviation

C) Tentative annotation of expected knowns

TL Target List based annotation for tentative ID based on HMDB 4.0 [3] derived urine metabolites

ID criteria	m/z	Isotopic pattern	RT	CCS	MS/MS
Target List	✓	✓		Prediction - CCSPredict Pro	*in-silico* fragmention - MetFrag

[3] 4306 compounds

InChI encoded structures enable:
- CCS prediction
- *in-silico* fragmentation

Name	Molecular Formula	InChI
1-Methylhistidine	$C_7H_{11}N_3O_2$	InChI=1S/C7H11N3O2/c1-10-3-5(9-4-10)2-6(8)7(11)12/h3-4,6H,2,8H2,1H3,(H,11,12)/t6-/m0/s1
1,3-Diaminopropane	$C_3H_{10}N_2$	InChI=1S/C3H10N2/c4-2-1-3-5/h1-3,4-5H2
2-Ketobutyric acid	$C_4H_6O_3$	InChI=1S/C4H6O3/c1-2-3(5)4(6)7/h2H2,1H3,(H,6,7)
2-Hydroxybutyric acid	$C_4H_8O_3$	InChI=1S/C4H8O3/c1-2-3(5)4(6)7/h3,5H,2H2,1H3,(H,6,7)/t3-/m1/s1
⋮		

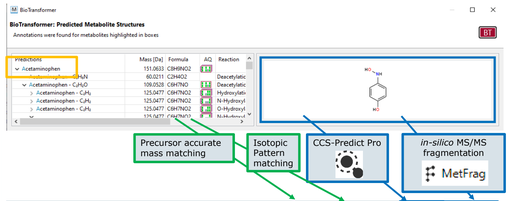
Glycylleucine $C_8H_{16}N_2O_3$ InChI=1S/C8H16N2O3/c1-5(2)3-6(8

Precursor accurate mass matching | Isotopic Pattern matching | CCS-Predict Pro | *in-silico* MS/MS fragmentation - MetFrag

Name	Molecular Formula	AQ	AQ Details	Δm/z [ppm]	mSigma	ΔRT	ΔCCS [%]	MS/MS score
Glycylleucine	$C_8H_{16}N_2O_3$			0.459	6.6		-0.1	801.9
Cytosine	$C_4H_5N_3O$			-1.972	6.4		-0.7	934.3
3-Methylglutarylcarnitine	$C_{13}H_{23}NO_6$			0.086	14.2		0.3	828.6
Pimelylcarnitine	$C_{14}H_{25}NO_6$			-0.194	14.4		0.5	868.5
Indolelactic acid	$C_{11}H_{11}NO_3$			0.551	10.0		0.3	844.0
⋮								

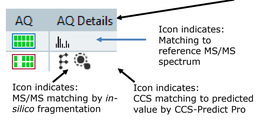
D) Tentative annotation of BioTransformer[4] predicted knowns

BT MetaboScape enables triggering of local BioTranformer based prediction of metabolites based on target structures. Visualization allows for the review of the generated metabolism tree and automatically matched compounds in the Feature Table:



Precursor accurate mass matching | Isotopic Pattern matching | CCS-Predict Pro | *in-silico* MS/MS fragmentation - MetFrag

Name	Molecular Formula	AQ	AQ Details	Δm/z [ppm]	mSigma	ΔRT	ΔCCS [%]	MS/MS score
Acetaminophen	$C_8H_9NO_2$			1.394	6.1	0.14	0.0	902.2
Acetaminophen + CH_2O	$C_9H_{11}NO_3$			-1.088	6.1		0.5	1000.0
Acetaminophen + $C_4H_4O_4$	$C_{12}H_{13}NO_6$			0.590	2.2		-0.3	1000.0
Acetaminophen + $C_4H_6O_3$	$C_{12}H_{15}NO_5$			-0.068	18.8		0.3	968.0
Acetaminophen + O_3S (1)	$C_8H_9NO_5S$			0.485	15.9		-2.7	987.3

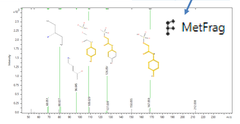


Icon indicates: Matching to reference MS/MS spectrum

Icon indicates: MS/MS matching by *in-silico* fragmentation

Icon indicates: CCS matching to predicted value by CCS-Predict Pro

Combined, AQ and AQ Details enable to easily assess confidence in ID and subsequent custom reporting of ID confidence levels



Visual inspection of MetFrag generated *in-silico* fragment structures enables to readily investigate tentative structure annotations.

Results

A) 4D-Metabolomics™ data acquisition of a human urine sample by LC-TIMS using PASEF enables MS/MS acquisition at >120 Hz, increasing the depth of coverage for all small molecules. This increased MS/MS coverage, in this study >80%, increases confidence levels by the availability of detected features that contain MS/MS spectra. TIMS intrinsically generates ion mobility information which is automatically recalibrated and transformed into CCS values in MetaboScape® by the T-ReX® 4D feature extraction algorithm.

B) Annotating the urine metabolite extract with a target list of known compounds derived from CCS Compendium [2] and appending retention time information and MS/MS spectra from several complementary reference spectral libraries permitted annotation of metabolites with highest confidence. 5 criteria could be matched to reference values (m/z, isotopic pattern, RT, MS/MS, CCS). AQ scoring enabled to readily check confidences for each automatic annotation made. Structures assigned by this annotation (example shown for acetaminophen) can serve as a starting point for metabolite prediction (see D).

C) Annotating the data set with a Target List of compounds reported to be present in urine (derived from HMDB 4.0; [3]) enabled tentative annotation of further metabolites. Note: This Target List could be appended with RTs, MS/MS spectra and/or CCS values from reference repositories as highlighted in B. In case these are not readily available, MetaboScape can perform automatic CCS prediction and MS/MS matching based on InChi encoded structures. This is based on the novel CCS-Predict Pro model and MetFrag [4,5], respectively.

D) Starting from the acetaminophen annotation (see B), triggering the BioTransformer [6] tool in MetaboScape for the prediction of metabolites enabled the additional assignment of several acetaminophen metabolites. Automatic *in-silico* fragmentation and CCS prediction resulted in high MS/MS matching scores and low CCS deviation supporting the tentative annotations for the drug metabolites. Manual investigation of MetFrag for annotated structures can help to substantiate these tentative assignments.
The workflows described in B, C and D can generate annotations based on reference MS/MS spectra matching or *in-silico* fragmentation and reference CCS values or predicted CCS values. The AQ Details column for each annotation highlights these differences with intuitive icons. The AQ score allows researchers to quickly assess their confidence in each annotation and in combination with the AQ Details icon to report ID level confidence suitable for individual studies, as recommended by Schymanski *et al.* [1].

Summary

The described workflows provide automated annotation routines across different confidence levels for known and predicted metabolites, building on trapped ion mobility separation and MS/MS spectra.

References

[1] https://doi.org/10.1021/es5002105
[2] https://doi.org/10.1039/C8SC04396E
[3] https://hmdb.ca/
[4] https://doi.org/10.1186/1471-2105-11-148
[5] https://doi.org/10.1186/s13321-016-0115-9
[6] https://doi.org/10.1186/s13321-018-0324-5

Note: HMDB and CCS Compendium are no Bruker products.

Conclusions

- MetaboScape provides users highest confidence in ID by matching up to **5 confidence criteria** including **reproducible CCS values**
- MetaboScapes novel **CCS-Predict Pro** model enables small molecule CCS prediction based on the molecular structures of target compounds
- **Automatic *in-silico* fragmentation and CCS prediction** (in case of lacking reference MS/MS spectra and / or CCS values) provides users **higher confidence in annotations**
- Reporting of **Annotation Quality Scores** and **Annotation Quality Details** enables researchers to readily **assess** and **report** their **ID level confidence**

4D-Metabolomics

For research use only. Not for use in Clinical diagnostic procedures.