Increasing confidence for compound identification by fragmentation database and *in silico* fragmentation comparison with LC-HRAM-MS-based non-targeted screening of complex matrices

Christoph Buchholz, Daniel Arndt, Christian Wachsmuth, Mark Bentley

PMI R&D, Philip Morris Products S.A., Quai Jeanrenaud 5, CH-2000 Neuchâtel, Switzerland

Overview

Liquid chromatography coupled to high-resolution accurate mass mass spectrometry (LC-HRAM-MS)-based non-targeted screening (NTS) applies accurate mass (AM), isotopic similarity, retention time (RT), and tandem mass spectrometry (MS^2) fragment spectra comparison for compound identification in cigarette smoke. However, the lack of first order MS^2 fragment spectra from reference compounds impedes unambiguous assignment of unknowns derived from 3R4F[1] reference cigarette smoke samples.

Computational approaches, including *in silico* fragmentation, are considered to be promising tools to fill this gap with the objective of increasing the identification confidence for unknown compounds, thereby minimizing the number of putative annotations in commercial and/or in-house AM and/or compound databases.

For this approach, a Q Exactive™ is the instrument of choice, as it delivers robust MS^2 HRAM data on a scan-by-scan basis. Data processing, which includes the library search and fragmentation pattern prediction, is performed using Progenesis QI™.

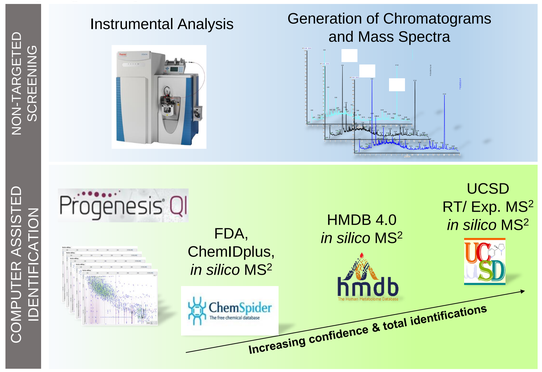

Figure 1. *Non-Targeted Screening with Thermo Q Exactive™ and integrated compound identification with Progenesis QI™*

Results

Compound Identification with in-house Databases

Accurate mass databases comprising compounds that occur in tobacco, such as UCSD (Unique Compounds & Spectra Database), are ideal in order to limit the compound proposals when searching for known small molecules in tobacco samples. If unknown compounds occur, a less matrix specific database needs to be queried resulting in increasing number of compound proposals based on accurate mass (***Fig 3.***).

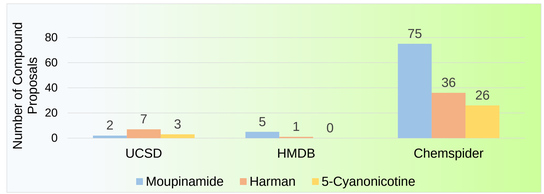

Figure 3. *Number of compound proposals based on accurate mass search for Moupinamide, Harman and 5-Cyanonicotine applying UCSD, HMDB and Chemspider data sources of ChemIDPlus and FDA*

Experimental databases for accurate mass, RT and MS^2 spectra provide the highest confidence for the identification of known compounds using a combined scoring of accurate mass, RT and MS^2 match. (***Fig 4.***)

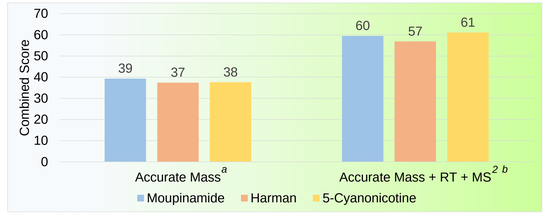

Figure 4. *Combined scoring of compounds queried with accurate mass only[a] and with a combination of accurate mass, retention time and MS^2 match retrieved from experimental databases[b]*

Compound Indentification with *in silico* Fragmentation

A similarity search of the *in silico*-predicted fragments with the experimental MS^2 spectra (**Figures 5, 6, 7,** and **8**) enhances the probability of the compound proposals. In addition, it is possible to differentiate between compound classes, as shown in **Table 1**. The molecular formula $C_{17}H_{26}N_2O$ has several Compound ID's in UCSD and HMDB. The AM and *in silico* fragmentation are included into the score of 51.5 for N-octanoylnornicotine, which is the highest-scoring candidate. The matched *in silico* fragments of the lower-scoring candidates are shown in **Figures 6, 7,** and **8**. N-octanoylnornicotine is not commercially available and had to be custom-synthesized. After custom synthesis, the RT was confirmed with an error of -0.03 min, and an MS^2 fragmentation score of 93.5 (**Figure 9**) was achieved. The score for N-octanoylnornicotine increased from 51.5 to 67.5 (**Table 1**).

Compound ID	Description	*AM, in silico Score*[a]	*AM, RT, MS2 Score*[b]
UCSD PMI0001863	**N-octanoylnornicotine**	**51.5**	**67.5**
UCSD PMI0006628	(S)-1-(6-methyl-1-oxoheptyl)-2-(3-pyridinyl)-Pyrrolidine	50.0	50.0
HMDB37992	Cyanidin 3-(diferuloylsophoroside) 5-glucoside	44.2	44.2
HMDB14441	Ropivacaine	43.2	43.2

Table 1. *[a]The score consisting of AM and in silico fragmentation enhances the probability of compound proposals. [b]The score including AM, RT, and MS^2 is retrieved from the injection of a reference material after custom synthesis.*

Figure 5. *In silico-predicted fragments for N-octanoylnornicotine. Fifteen matched in silico-predicted fragments are marked in red.*

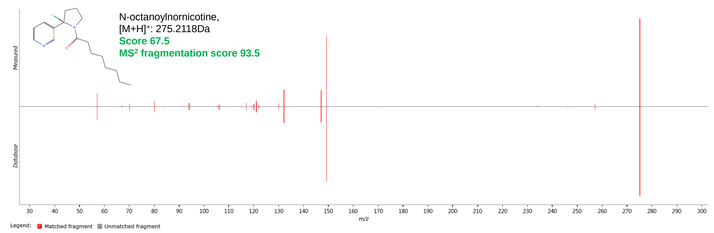

Figure 9. *Combined score of 67.5 including AM, RT, MS^2 fragments for N-octanoylnornicotine with custom-synthesized reference material. The MS^2 fragment score was 93.5.*

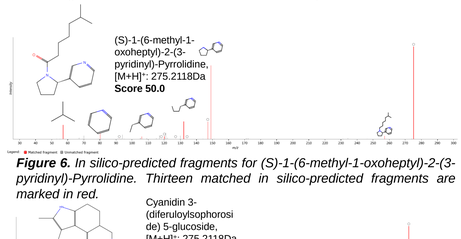

Figure 6. *In silico-predicted fragments for (S)-1-(6-methyl-1-oxoheptyl)-2-(3-pyridinyl)-Pyrrolidine. Thirteen matched in silico-predicted fragments are marked in red.*

Figure 7. *In silico-predicted fragments for* Cyanidin 3-(diferuloylsophoroside) 5-glucoside. *Twelve matched in silico-predicted fragments are marked in red.*

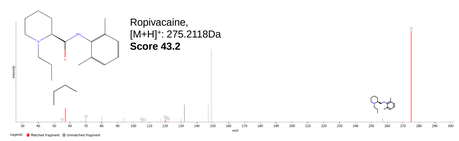

Figure 8. *In silico-predicted fragments for* Ropivacaine. *Nine matched in silico-predicted fragments are marked in red.*

Methods

A total of four methods (e.g., reversed phase (RP) heated electrospray ionization (HESI) positive, RP HESI negative, RP atmospheric pressure chemical ionization positive, hydrophilic interaction liquid chromatography HESI positive) were applied for an NTS of a 3R4F cigarette smoke sample. Data acquisition was performed using a Q Exactive™ Hybrid Quadrupole Orbitrap MS (Thermo Scientific, Germany) in connection with an Accela 1250 UHPLC pump. Data processing was performed with Progenesis QI™ (Nonlinear Dynamics, UK), with an in-built MetFrag[2] algorithm for enhanced *in silico* prediction of unknown compounds. Databases such as the Unique Compounds & Spectra Database (UCSD[3]), HMDB 4.0[4], and Chemspider with ChemIDplus and U.S. Food and Drug Administration data sources were queried simultaneously for compound screening based on AM and isotope similarity. An in-house RT and MS^2 database matched RTs and MS^2 spectra of known compounds. *In silico* predicted fragments were generated and matched against the acquired MS^2 fragment spectra.

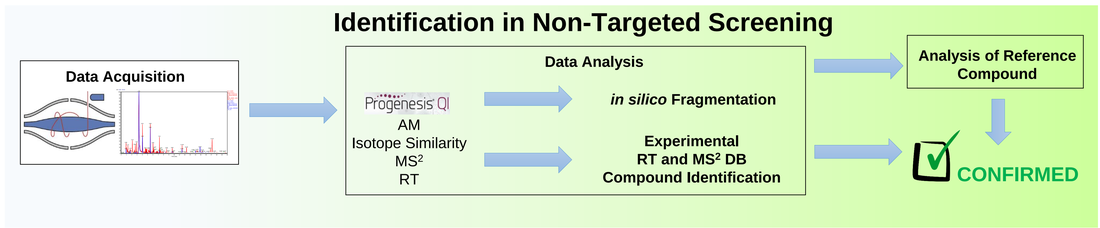

Figure 2. *Compound identification workflow. Identified compounds using in silico fragmentation proposals are reanalyzed with the respective reference compound, leading to the ultimate goal of compound confirmation.*

Conclusions

A workflow has been established to improve the identification of compounds derived from LC-HRAM-MS-based NTS using both full scan and MS^2

- The highest confidence and confirmation are achieved by matching acquired data with in-house RT and MS^2 spectra databases of reference compounds.
- It has been demonstrated that *in silico* fragmentation can minimize the number of proposals for compound identification and increase the confidence in the selected candidates.
- The highest-scoring *in silico* fragmented compound was custom synthesized, which successfully confirmed the RT and MS^2 spectrum for final compound confirmation and ultimate confidence.

References

[1] Roemer, E, et al., *Contributions to Tobacco Research,* **2012**, 25, 316.
[2] Wolf S., Schmidt S. Mueller-Hannemann M., Neumann S., *BMC Bioinformatics,* **2010**, 11, 148.
[3] Martin, E.; Monge, A.; Duret, J. A., et al., *J Cheminform,* **2012**, 4, 11.
[4] Wishart, D. S.; Feunang, Y. D.; Marcu, A., et al., *Nucleic Acids Res,* **2018**, 46, D608.

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

66th ASMS Conference on Mass Spectrometry and Allied Topics San Diego, CA, USA
June 3 – 7, 2018

Competing Financial Interest

The research described in this poster was sponsored by Philip Morris International