Genomics driven metabolomics novel strategies for the discovery and identification of secondary metabolites

Ries, M.

Chapter 6

Multiple stage fragmentation tree comparison enables detailed structure elucidation in direct infusion mass spectrometry based experiments

Based on

Marco I. Ries, Hazrat Ali, Arnold J.M. Driessen, Thomas Hankemeier, Theo Reijmers, Rob J. Vreeken
Multiple stage fragmentation tree comparison enables detailed structure elucidation in direct infusion mass spectrometry based experiments
In preparation for publication

Abstract

Direct infusion mass spectrometry (DI-MS) is a fast, reliable, sensitive and cost effective method which is particularly attractive when dealing with large sample sets. However, structure identification in direct infusion based experiments is still a major obstacle due to co-fragmenting interferences like isobaric and isomeric ions, limiting most applications to metabolite fingerprinting only.
Here we present the application of CMCP (Chemoinformatics supported MS^n Comparison Pipeline) for the identification of metabolites in direct infusion based experiments. We demonstrate the structure elucidation of secondary metabolites in a complex liquid culture of *Penicillium chrysogenum* using multiple stage fragmentation spectra from single precursor ions as well as with isobaric and isomeric interferences. Next to the structure elucidation with the sample compound being present in the database, structure identification based on similar structures is described enabling *de novo* structure elucidation in direct infusion based experiment beyond a classification of compounds into compound classes.
This study presents a potential method to identify unknown metabolites conveniently in DI, without additional instrumental configurations, extending DI to more than just a first line approach.

Introduction

With the prevalent application of metabolomics, several methods and techniques for the simultaneous analysis of a large number of compounds have been developed which provide detailed information about the metabolome of complex biological samples. Traditional approaches for the determination of metabolites by mass spectrometry use chromatographic separation like GC (Gas Chromatography), LC (Liquid Chromatography) or CE (Capillary Electrophoresis) before metabolite detection, thus generating qualitative or quantitative information on individual analytes (Dettmer, et al., 2007). These methods require laborious and time consuming sample preparation procedures and significant analysis time due to chromatographic separation. Furthermore, separation and detection of metabolites is heavily depending on the chemical nature of the analytes and the used chromatographic method. In general, a combination of various analytical techniques like HILIC, RP or Ion-pairing, targeting different classes of metabolites, is necessary as no single analytical platform allows a comprehensive analysis of the metabolome.
An alternative high-throughput approach to capture information related to the total metabolite content is the use of direct infusion mass spectrometry (DI-MS) (Boernsen, et al., 2005; Goodacre, et al., 2002; Koulman, et al., 2007) which renounces the limiting separation step before sample ionization, allowing the fast detection of various metabolite classes. During the ionization process of electrospray ionization (ESI), the most common method used for ionization in mass spectrometry based experiments, analytes become charged by the loss or gain of a proton, or other adducts. As minimal fragmentation takes place during ionization, the measured mass of an analyte, recorded with high mass accuracy and resolution, is expected to be close to the anticipated mass of a specific metabolite recorded in a database, potentially allowing a direct putative identification (Mungur, et al., 2005; Nakamura, et al., 2007; Raterink, et al., 2013). Several databases containing metabolite information are available in order to facilitate putative annotation of accurate mass signals (e.g. PubChem (Li, et al., 2010), HMDB (Wishart, et al., 2009), KEGG (Kanehisa, et al., 2012), KNapSAcK (Afendi, et al., 2012)). However, an accurate mass does not give any structural information beyond the molecular formula, limiting most direct infusion experiments to screening purposes only (Draper, et al., 2013). Using MS/MS, individual precursor ions can be selected and fragmented, resulting in specific fragment ions which can help to elucidate the structure of the molecule. The interpretation of MS/MS experiments is challenging as isobaric ions (same nominal mass) and isomeric ions (same chemical composition) are fragmented simultaneously, resulting in MS^2 spectra with mixed fragments of various precursor ions (Wichitnithad, et al., 2010). Although newer generation mass spectrometers substantially reduce co-fragmentations with isolation widths as narrow as 0.5 *m/z* (mass-to-charge ratio) (Savitski, et al., 2011), co-fragmentation of interferences can't be completely avoided leaving structure identification in direct infusion a major challenge. Therefore, more time consuming methods are currently required for a full structure elu-

cidation involving chromatographic separation and MS or NMR based approaches which in return, compromise the advantages of DI experiments. Especially finding back the target molecules, obtained from DI screening, in chromatographic based experiments is challenging as the physiochemical properties of the targeted analyte are mostly unknown.
In contrast to MS/MS approaches, multiple stage mass spectrometry (MS^n) enables a further fragmentation of fragments generating so called fragmentation trees. As solely selected ions are further fragmented, insight into multidimensional precursor-parent ion relationships is obtained. With the chemical structure of the precursor coded in its fragments and neutral-losses, structural information of the precursor ion can be deduced (Cui, et al., 1999; Kang, et al., 2007; Rochfort, et al., 2008). Recently, our group developed CMCP for the structure elucidation of compounds using solely multiple stage fragmentation tree comparison (**Chapter 5**). By comparing fragment and neutral-loss trees, compounds with similar fragmentation mechanisms can be extracted from a MS^n database (Kasper, et al., 2012; Rojas-Cherto, et al., 2012). Based on the identified shared fragmentation mechanisms, chemical structures of unknowns can be deduced without them being present in a database.

Here we show the enormous potential of CMCP for the structure elucidation of compounds from complex analyte mixtures using direct infusion. To demonstrate the capabilities of our approach to identify compounds from fragmentation data of one population of structurally identical precursors as well as from co-fragmented isobaric and isomeric ions, the structure elucidation of various complex metabolites from liquid cultures of the fungus *Penicillium chrysogenum* is described.

Experimental procedures

Material and Chemicals
3-me-7-(3-methylbenzyl)-8-((1-phenylethyl)amino)-3,7-dihydro-1H-purine-2,6-dione (**2**) was purchased from Sigma-Aldrich.

*Preparation of working and stock solutions of **1** and **2***
Stock solutions consist of 1 µg/mL analyte in ethanol. A working solution was prepared by mixing the stock solutions of compound **1** and **2** in a 1:5 ratio. Right before acquisition 10 µL working solution was mixed with 5 µL isopropanol and 10 µL water containing 0.1 % formic acid in a 384 well plate (Eppendorf, Hamburg, Germany).

Strain and culture conditions
P. chrysogenum strain DS54555, which lacks the *Ku70* and penicillin cluster genes was kindly supplied by DSM Anti-infective (Delft, The Netherlands). Cells were grown on SMP medium (glucose, 5.0 g/L; lactose, 75 g/L; urea, 4.0 g/L; Na_2SO_4, 4.0 g/L; CH_3COONH_4, 5.0 g/L; K_2HPO_4, 2.12 g/L; KH_2PO_4, 5.1 g/L) using a shaking incubator at 200 rpm for 168 hours at 25°C.

Sample preparation
To 50 µL of a thawed fermentation broth 8 µL internal standard mixture containing 855 nmol/mL ranitidine, 657 nmol/mL reserpine and 1144 nmol/mL ampicillin was added. Subsequently, 230 µL of methanol was added for protein precipitation and vortexed for 10 minutes. The sample was then centrifuged at 14,000 g for 10 minutes at 10°C. 100 µL supernatant was transferred to an Eppendorf vial and evaporated for 30 minutes in a Thermo-Speedvac (Thermo Scientific, San Jose, CA). The dried sample was resolved in 100 µL water containing 2 % acetonitrile and vortexed for 10 minutes.

Fragmentation tree acquisition, processing and database comparison
Fragmentation tree acquisition was carried out as previously described (**Chapter 5**) in positive ion mode on a LTQ-Orbitrap-XL mass spectrometer (Thermo Fisher Scientific, Waltham, MA) using an Advion Nanomate (Advion, Ithaca, NY) for nano-electrospray with following settings: 1,5kV source voltage, 120°C capillary temperature, 5V capillary voltage, 70V tube lens voltage and 35 % normalized collision energy for CID fragmentation. A minimum of three repetitions of the complete fragmentation tree were acquired.
Assigned fragmentation trees were created from acquired Thermo Xcalibur files using a macro programmed in Excel (Microsoft, Redmond, WA). Finally, all processed trees were compared to an in-house fungal metabolite MS^n database for determination of similar fragments and losses (Rojas-Cherto, et al., 2012).

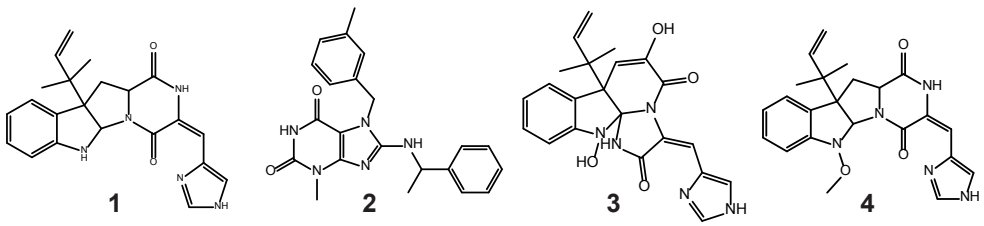

Figure 1. Structures of discussed compounds

Results and Discussion

To demonstrate the capability of CMCP to confidently elucidate the identity of unknown compounds in DI based experiments despite a co-fragmentation of interferences, the structure identification of the four analytes roquefortine C (**1**, $C_{22}H_{23}N_5O_2$, calc. $[M+H]^+$ = 390.1924), 3-me-7-(3-methylbenzyl)-8-((1-phenylethyl) amino)-3,7-dihydro-1H-purine-2,6-dione (**2**, $C_{22}H_{23}N_5O_2$, calc. $[M+H]^+$ = 390.1924) , glandicoline B (**3**, $C_{22}H_{21}N_5O_4$, calc. $[M+H]^+$ = 420.1666) and roquefortine F (**4**, $C_{23}H_{25}N_5O_3$, calc. $[M+H]^+$ = 420.2030) (Figure 1) is described under various conditions. First, structure identification based on identical database entries is shown for single, isomeric and isobaric ions, representing identity searches in DI experiments. To demonstrate *de novo* structure identification, in which the mass spectral data of the analyte is not represented in a database, the structure elucidation of **1** is shown using complementary fragment and neutral-loss tree comparison.

Fragmentation tree generation and database search

Multiple stage fragmentation spectra for each compound were acquired using direct infusion as described in the Materials and Methods section. Due to the gentle ionization of ESI, little fragmentation of precursor ions was observed. After elemental compositions were assigned to all fitting ions present in the acquired fragmentation tree using a fixed precursor composition, fragment and neutral-loss trees were automatically generated. The resulting trees were compared to an in-house MS^n database searching for similar mass fragments and neutral-losses. Obtained database entries were ranked in a descending order according their similarity represented by the Tanimoto coefficient (Flower, 1998).

Identity search of single ion trees

The database query of the fragment tree of compound **1**, which did not show additional abundant ions present in the same isolation range of MS^1 prior fragmentation, returned the database entry roquefortine C as most similar hit. As all fragments and neutral-losses of compound **1** could be found present and similarly arranged in the database entry, an identical structure for both compounds was indicated (Figure 2). This was supported by comparing their individual multiple stage fragmentation spectra showing almost identical mass spectra on every level (data not shown). Therefore, **1** was tentatively identified as roquefortine C which was ultimately confirmed using HPLC-MS/MS, comparing **1** and a standard of roquefortine C.

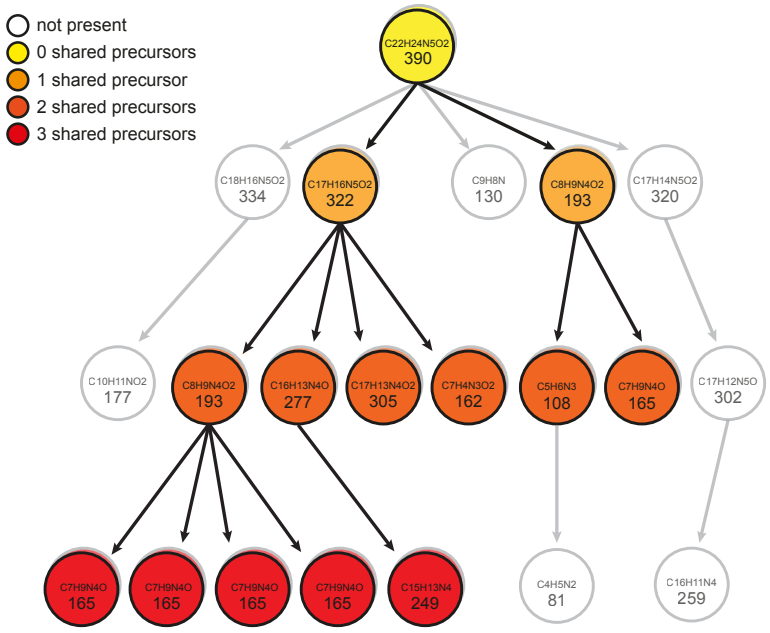

Figure 2. Fragment tree comparison without co-fragmented ions

Superimposed fragment trees of **1** (top), acquired from a biological samples using DI-MS, and partial fragmentation tree of database entry roquefortine C (back) after fragment tree comparison. All fragments of **1** could be found similarly arranged in the database entry (colored nodes) whereas not all fragments of the database entry could be found in the fragment tree of **1** (white nodes). Differences between both trees are due to different precursor intensities.

The difference in depth and width between the fragment tree of **1** and roquefortine C results from the different precursor intensities during fragmentation tree acquisition. The tree of **1** was acquired using a complex biological sample with a low concentration of **1**, whereas the fragmentation tree of roquefortine C in the reference database was acquired with a highly concentrated standard resulting in a much richer fragmentation.

Identity searches of isomeric ion trees

We next investigated the possibility of achieving unambiguous identification of co-fragmented isomeric precursor ions using CMCP by spiking the isomer **2** to a solution of **1**. With an identical mass-over-charge ratio in MS^1, the two isomers **1** and **2** with *m/z* 390.1926 ($C_{22}H_{23}N_5O_2$, calc. $[M+H]^+$ = 390.1924) are simultaneously fragmented resulting in one single fragmentation tree containing fragments from both compounds (Figure 3 and Figure 4A). Due to an identical elemental composition of their precursor ions in MS^1, all fragments originating from a fragmentation of **1** and **2** could be assigned with their correct elemental composition when processed with the chemical formula of protonated **1**. This resulted in the incorporation of all fragments of **2** into the fragment tree of **1** (Figure 4B). Using CMCP, the database query returned the protonated database entry roquefortine C ($C_{22}H_{23}N_5O_2$, $[M+H]^+$ = 390.1924) as most similar hit to **1**, sharing a common precursor ion next to several subtrees initiated in MS^2, thus indicating that **1** and roquefortine C are

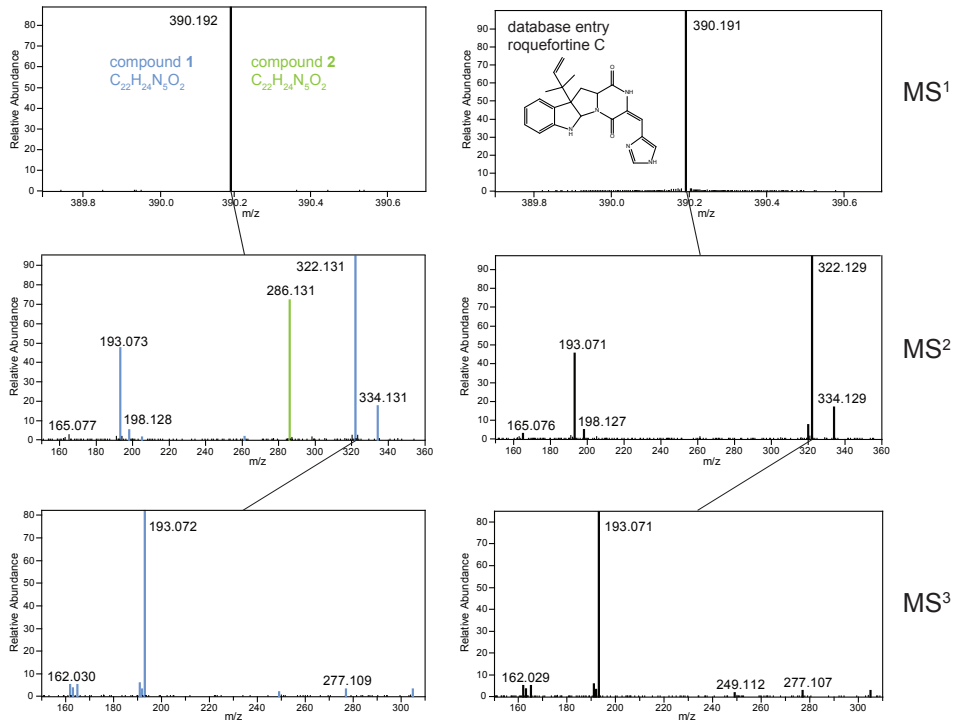

Figure 3. Multiple stage fragmentation spectra of the co-fragmented isomers **1** and **2** (left), and the database entry roquefortine C (right). Fragments originating from compound **1** are colored blue whereas fragments of **2** are colored green.

structurally related (Figure 4C). This was further supported by highly similar MS^2 spectra in which all fragments of roquefortine C were present with a similar mass-to-charge ratios and relative intensities in the MS^2 spectrum of co-fragmented **1** (Figure 3). In contrast, fragments originating from a co-fragmentation of **2**, like *m/z* 286.131 were highly abundant in the MS^2 of **1** but not present in the MS^2 spectra of the database entry roquefortine C, resulting in ambiguity about the structure of **1**. With the automatic multiple stage fragmentation spectra acquisition in CMCP, product ions in MS^2 were fragmented further yielding fragmentation spectra of various fragmentation stages, corresponding to a fragmentation of **1** which were free of interfering ions and completely identical to their analogue in the database entry roquefortine C. So shows the MS^3 spectra of *m/z* 322.131 in **1** and roquefortine C similar fragments in terms of mass-to-charge ratios and relative intensities (Figure 3, Figure 4C) leading to the conclusion that **1** and roquefortine C are structurally identical. In addition, fragments in the fragment tree of **1**, which were not found to be shared by the database entry roquefortine C, have to originate from a fragmentation of at least one additional isomer, namely compound **2** (Figure 4C).

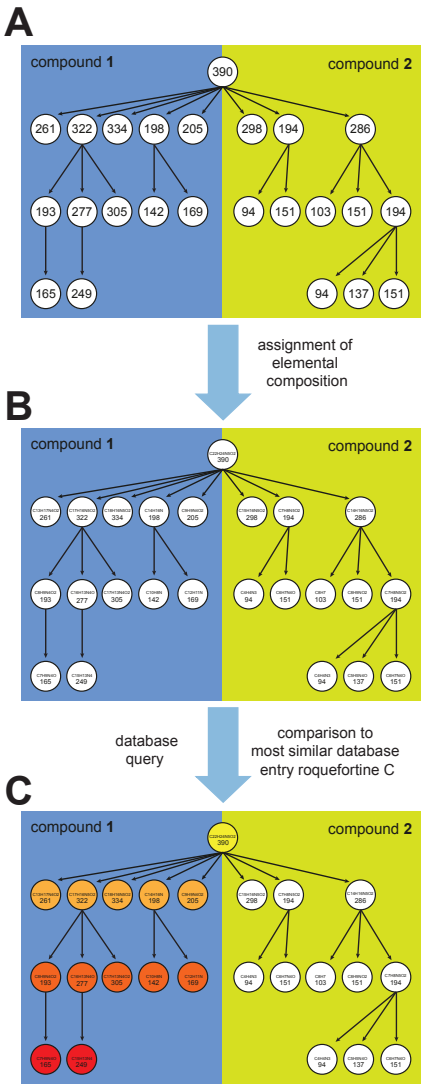

Figure 4. Structure elucidation of co-fragmented isomers. Fragment tree comparison for the co-fragmented isomers 1 and 2. As both compounds have the same elemental composition, all fragments of 2 were assigned with a correct elemental composition when processed with the precursor formula of protonated 1. The database query returned the database entry roquefortine C as most similar hit, with all fragments of 1 being present and similarly arranged.

In addition, as the complete spectral information of co-fragmented **2** is contained in the acquired fragmentation tree of **1**, the identification of **2** can be performed accordingly from the same multiple stage fragmentation data. Using the second most similar database entry 3-me-7-(3-methylbenzyl)-8-((1-phenylethyl) amino)-3,7-dihydro-1H-purine-2,6-dione from the database query, which shares solely the remaining fragments with the fragment tree of **1** and **2**, the structure of **2** could be determined.

These results demonstrate the confident identification of analytes in direct infusion using CMCP despite a co-fragmentation of isomeric ions.

Identity searches of isobaric ion trees

When analyzed separately, compound **3** with *m/z* 420.1659 ($C_{22}H_{21}N_5O_4$, cal. $[M+H]^+$ = 420.1666) could be unambiguously identified based on its fragmentation tree, following the strategy described for the Identity search of compound **1**. However, when analyzed in the liquid culture of *P. chrysogenum*, a compound with a very similar mass-over-charge ratio of 420.2024 (**4**, $C_{23}H_{25}N_5O_3$, calc. $[M+H]^+$ = 420.2030) was observed.

Although clearly distinguishable in MS^1, due to the high-resolution of the Orbitrap, compound **3** and the interfering isobaric ion of **4** were isolated together and fragmented simultaneously due to the insufficient resolving power available in the isolation step. This resulted in an identical fragmentation tree with identical MS^n spectra for both compounds, containing fragments originating from a fragmentation of **3** and **4** (Figure 5). By processing the acquired multiple stage fragmentation spectra with the chemical formula of protonated **3**, ions originating from a fragmentation of **3** were assigned with their correct elemental composition and incorporated into the fragmentation tree. In contrast, fragments originating from a fragmentation of **4** were mainly rejected as their elemental composition was not consistent with the chemical formula of **3** (Figure 6A and B). So is the fragment with *m/z* 389 ($C_{22}H_{23}N_5O_2$) of **4** and its complete subtree, not present in the fragmentation tree of **3** as its corresponding elemental composition contains two additional hydrogens com-

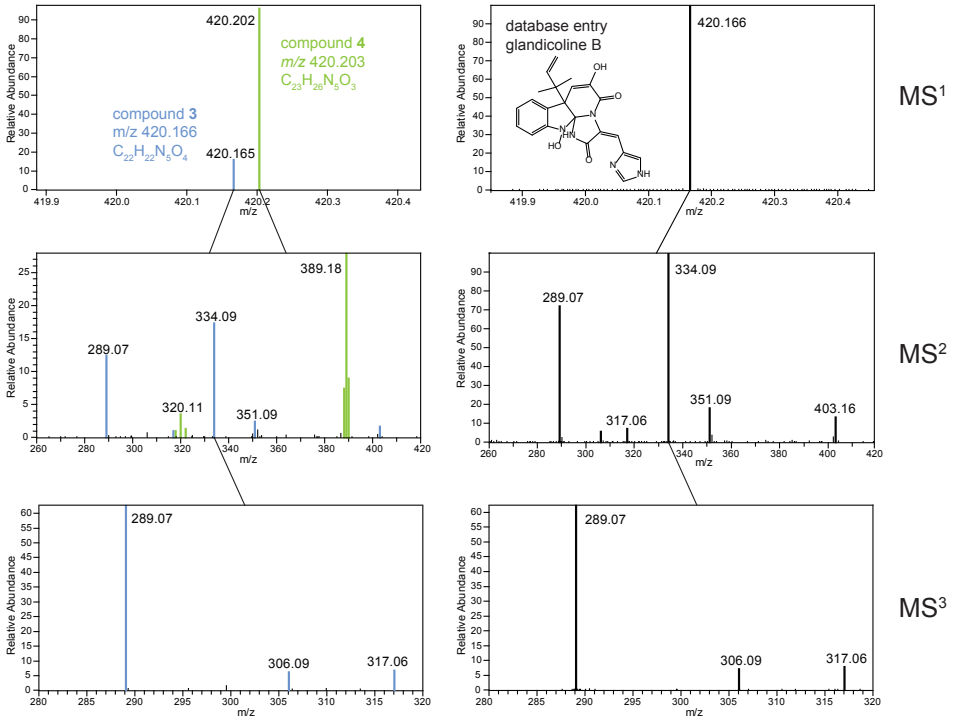

Figure 5. Multiple stage fragmentation spectra of the co-fragmented isobars **3** and **4**, acquired from a biological sample using DI-MS (left), and the database entry glandicoline B (right). Fragments originating from compound **3** are colored blue whereas fragments of **4** are colored green.

pared to protonated **3**. However, due to the relatively similar elemental composition of **3** and **4**, few fragments originating from a fragmentation of **4** could be still assigned, like *m/z* 320 ($C_{17}H_{14}N_5O_2$), *m/z* 350 ($C_{18}H_{16}N_5O_3$) and *m/z* 388 ($C_{22}H_{22}N_5O_2$). The subsequent database query of **3** returned highest similarity to the database entry glandicoline B ($C_{22}H_{21}N_5O_4$, $[M+H]^+$ = 420.16663) which shares next to similarly arranged fragments and neutral-losses also its precursor ion with **3** (Figure 5 and Figure 6C). MS^2 fragments, present in both compounds, showed similar subtrees, indicating that **3** and glandicoline B share a similar (sub)structure. This was supported by comparing their individual multiple stage fragmentation spectra showing almost identical relative intensities of fragments shared by **3** and glandicoline B (Figure 5). Therefore, it is concluded that remaining fragments in the fragment tree of **3**, which were not found shared by the database entry glandicoline B, correspond to a fragmentation of **4** (Figure 6C). The structure of **3** was tentatively identified as glandicoline B and ultimately confirmed using HPLC-MS/MS comparing **3** and a chemical standard of glandicoline B.

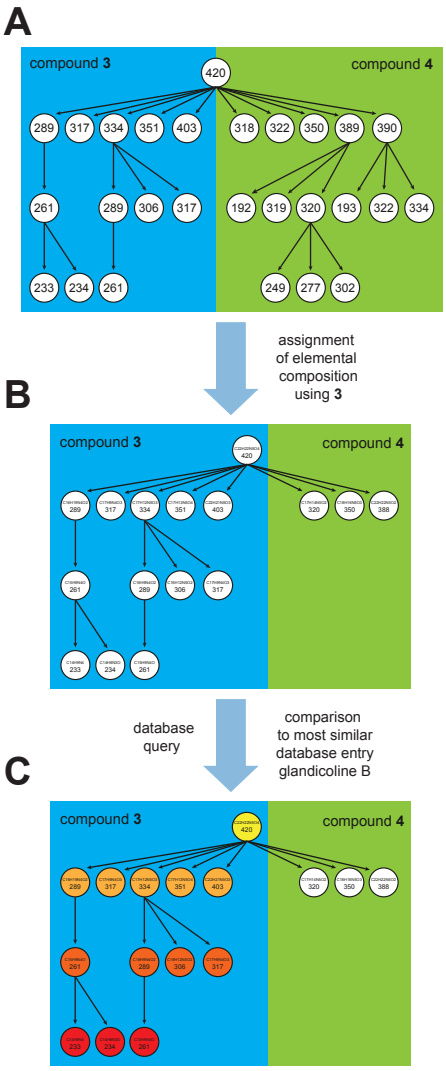

Figure 6. Structure elucidation of co-fragmented isobars
Fragment tree comparison of the co-fragmented isobars **3** and **4**. First, an unassigned fragmentation tree containing fragments of compound **3** and **4** was obtained. During the assignment of elemental compositions to fragments using the precursor formula of **3**, several fragments originating from **4** were rejected due to inconsistency. Last, the database query returned the database entry glandicoline B as most similar entry with all fragments of **3** being present and similarly arranged. The complete fragment trees of **3** and **4** can be found in the Supplement.

As **3** and **4** were fragmented simultaneously, the structural information of **4** is contained in the same multiple stage fragmentation spectra which were used to identify **3**. By processing the fragmentation tree of co-fragmented **3** and **4** with the elemental composition of **4**, instead of **3**, the fragment and neutral-loss tree of **4** were obtained (data not shown). Their comparison to the MS^n database returned roquefortine F as most similar database entry sharing almost all fragments and neutral-losses with **4**. Ultimately, **4** was tentatively identified as roquefortine F which was confirmed using NMR.

The structure identification of co-fragmented isobaric ions is comparable to the identification of compounds from single or isomeric ions, depending on

the elemental composition and fragmentation of the compounds. In case none of the fragments in MS^2 of the interfering isobaric compound are consistent with the precursor formula of the targeted analyte, a fragment tree without interfering fragments is obtained for the analyte resembling the identification of a single fragmented compound. If however, both isobars share a similar elemental composition, fragments and neutral-losses of the co-fragmented isobaric ions can be incorporated into the processed trees of the analyte, resembling the fragment and neutral-loss trees of co-fragmented isomers. Therefore, similar strategies for their identification apply.
In contrast to the identification of one population of structurally identical ions or co-fragmenting isomeric ions, the presence of isobaric ions is immediately recognized as such, as more than one precursor ion in MS^1 can be observed if the resolution of the mass spectrometer is sufficient enough.

De novo structure identification
So far, the structure elucidation shown was achieved using an identical standard present in the MS^n database. However, an identical database entry is not always available, especially when completely new compounds with novel structures were discovered. To demonstrate the *de novo* structure identification of single fragmented **1**, using not an identical but similar database entry, the database entry of **1** (roquefortine C) was removed from the database. Without an identical entry present in the database, the database query of **1** with *m/z* 390 ($C_{22}H_{23}N_5O_2$, calc. $[M+H]^+ = 390.1924$), returned highest similarity to the database entry dehydrohistidyltryptophanyldiketopiperazine (DHTD) with *m/z* 322 ($C_{17}H_{15}N_5O_2$, calc. $[M+H]^+ = 322.1299$), which shares almost its complete fragment and neutral-loss trees with **1**, as well as the database entry roquefortine D with *m/z* 392 ($C_{22}H_{25}N_5O_2$, calc. $[M+H]^+ = 392.2081$) which shares primarily a consecutive neutral-loss path (Figure 7A and 7B). As both database entries have a different elemental composition as **1**, it can be immediately concluded that a similar substructure rather than an identical structure is shared. Based on their shared fragments and neutral-losses, similar fragmentation mechanisms could be identified and a (sub)structure of **1** deduced (**Chapter 5**). As all fragments and neutral-losses in the fragment and neutral-loss tree of **1** originate from a fragmentation of **1**, the remaining unshared fragments can be used to validate the deduced structure for consistency, providing additional confidence for the proposed structure.
Different to the fragmentation tree of single fragmented **1**, in which all fragments originate from a fragmentation of **1**, the fragmentation tree of the co-fragmented isomers **1** and **2** contains additional fragments and neutral-losses from a fragmentation of **2** (Figure 7A and 7B). Although the same fragments and neutral-losses are shared with the database entries DHTD and roquefortine C, as shown for **1** without a co-fragmented interfering ion, a lower similarity value is obtained due to the additional unshared fragments and neutral-losses of **2**. Based on the shared nodes and the known structure of the database entry, a common fragmentation mechanism could be identified and a (sub)structure deduced for **1** (**Chapter 5**). In contrast to the previous identification, not all unshared fragments and neutral-losses can now be used to support the tentatively identified structure of **1** as they might originate from a fragmentation of **2**. However, a complete inconsistency with the tentatively

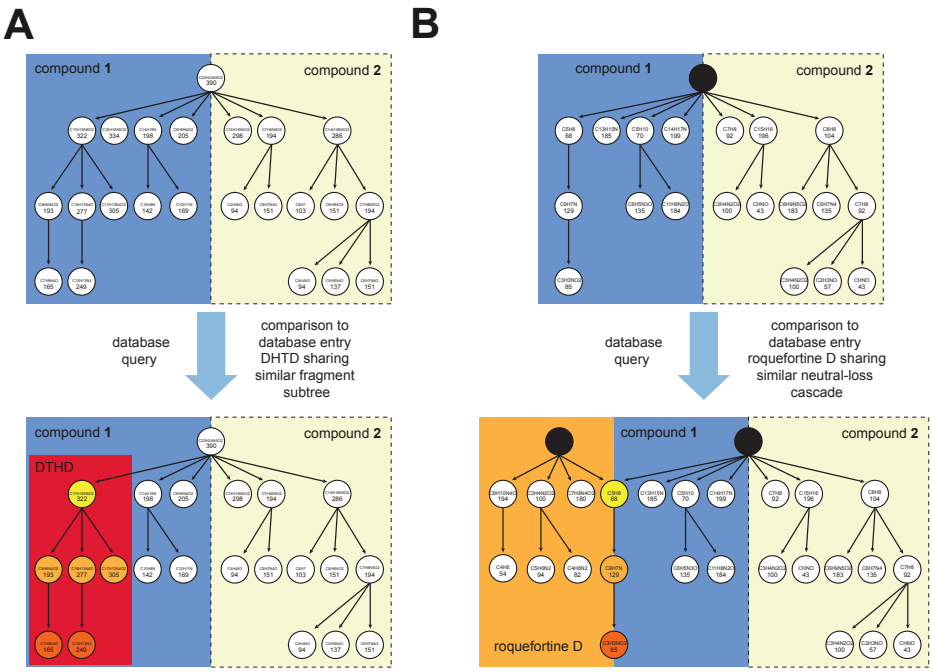

Figure 7. Schematic *de novo* identification of compound **1**, co-fragmented with the isomer **2**, using fragment tree comparison to the database entry DHTD (A) and neutral-loss tree comparison to the database entry roquefortine D (B). By removing the fragments and neutral-losses of **2**, indicated by a dashed square, the *de novo* identification of **1** without a co-fragmented ion can be deduced.

identified structure of **1** can give an indication that structurally different fragments are present most likely originating from a co-fragmentation of an isomeric precursor ion.

Comparable to the identity search of co-fragmented isobaric ions, shows the *de novo* structure elucidation of co-fragmented isobars either more similarity to the *de novo* identification of compounds from single ions or isomeric ions, depending on the elemental composition and fragmentation of the involved compounds. Although the presence of isobaric ions can be clearly recognized based on the presence of more than one precursor ion in MS^1, it can pose a similar challenge as the *de novo* identification of co-fragmented isomers if all fragments of the interfering isobar can be assigned with a valid elemental composition when processing with the molecular formula of the analyte. Therefore, similar strategies for their identification apply.

Conclusions

As a preceding separation step is absent in DI, all ions originating from a particular sample are, in general, measured at the same time in the mass spectrometer. Depending on the composition of the sample, unwanted ions can interfere with the compound of interest in form of isobaric or isomeric ions, resulting in their simultaneous fragmentation hampering structural identification. Different to MS/MS

approaches, fragment ions of an analyte can be structurally characterized using the Chemoinformatics supported MS^n Comparison Pipeline (CMCP) due to consecutive isolation and fragmentation allowing their differentiation from ions originating from co-fragmenting isobaric or isomeric precursor ions. With the subsequent generation of fragment and neutral loss trees from the acquired multiple stage fragmentation spectra, sub-trees from co-fragmented isobars can be confidently removed resulting in fragmentation data solely originating from the analyte. In cases where the sample compound is present in the database, a fully automated identification is possible despite co-fragmented interferences enabling detailed structure elucidation in direct-infusion based experiments. This makes CMCP a valuable tool as also newer generation mass spectrometer will not be able to completely avoid a co-fragmentation of isomeric and isobaric ions.
Furthermore, as CMCP is searching for common fragments and neutral-losses, also database entries which don't fragment in an identical, but similar fashion are returned. With the structural elucidation of shared fragments, deduced from the known structure of the database entry, confident *de novo* identification of unknown molecular (sub)structures can be achieved, even without the compound being present in the database. Therefore, less comprehensive databases are required to cover a large chemical space as not for every analyte an identical standard needs to be present in the database.
With the facile integration into already existing profiling pipelines, structural information can be directly obtained from DI experiments, offering more than just a 'first pass' screening without the immanent need to switch to another platform.

Acknowledgements

This project was supported by the Perspective Genbiotics program subsidized by the Stichting voor de Technische Wetenschappen (STW) and (co)financed by the Netherlands Metabolomics Centre (NMC) which is part of the Netherlands Genomics Initiative/Netherlands Organization for Scientific Research.

References

Afendi, F.M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K., Ikeda, S., Takahashi, H., Altaf-Ul-Amin, M., Darusman, L.K., et al. (2012). KNApSAcK family databases: integrated metabolite-plant species databases for multifaceted plant research. Plant Cell Physiol 53, e1.

Boernsen, K.O., Gatzek, S., and Imbert, G. (2005). Controlled protein precipitation in combination with chip-based nanospray infusion mass spectrometry. An approach for metabolomics profiling of plasma. Analytical chemistry 77, 7255-7264.

Cui, M., Sun, W., Song, F., Liu, Z., and Liu, S. (1999). Multi-stage mass spectrometric studies of triterpenoid saponins in crude extracts from Acanthopanax senticosus harms. Rapid communications in mass spectrometry : RCM 13, 873-879.

Dettmer, K., Aronov, P.A., and Hammock, B.D. (2007). Mass spectrometry-based metabolomics. Mass Spectrom Rev 26, 51-78.

Draper, J., Lloyd, A.J., Goodacre, R., and Beckmann, M. (2013). Flow infusion electrospray ionisation mass spectrometry for high throughput, non-targeted metabolite fingerprinting: a review. Metabolomics 9, 4-29.

Flower, D.R. (1998). On the Properties of Bit String-Based Measures of Chemical Similarity. J. Chem. Inf. Comput. Sci 38, 379-386.

Goodacre, R., Vaidyanathan, S., Bianchi, G., and Kell, D.B. (2002). Metabolic profiling using direct infusion electrospray ionisation mass spectrometry for the characterisation of olive oils. Analyst 127, 1457-1462.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 40, D109-114.

Kang, J., Hick, L.A., and Price, W.E. (2007). A fragmentation study of isoflavones in negative electrospray ionization by MSn ion trap mass spectrometry and triple quadrupole mass spectrometry. Rapid Commun Mass Spectrom 21, 857-868.

Kasper, P.T., Rojas-Cherto, M., Mistrik, R., Reijmers, T., Hankemeier, T., and Vreeken, R.J. (2012). Fragmentation trees for the structural characterisation of metabolites. Rapid Commun Mass Spectrom 26, 2275-2286.

Koulman, A., Tapper, B.A., Fraser, K., Cao, M., Lane, G.A., and Rasmussen, S. (2007). High-throughput direct-infusion ion trap mass spectrometry: a new method for metabolomics. Rapid Commun Mass Spectrom 21, 421-428.

Li, Q., Cheng, T., Wang, Y., and Bryant, S.H. (2010). PubChem as a public resource for drug discovery. Drug Discov Today 15, 1052-1057.

Mungur, R., Glass, A.D., Goodenow, D.B., and Lightfoot, D.A. (2005). Metabolite fingerprinting in transgenic Nicotiana tabacum altered by the Escherichia coli glutamate dehydrogenase gene. J Biomed Biotechnol 2005, 198-214.

Nakamura, Y., Kimura, A., Saga, H., Oikawa, A., Shinbo, Y., Kai, K., Sakurai, N., Suzuki, H., Kitayama, M., Shibata, D., et al. (2007). Differential metabolomics unraveling light/dark regulation of metabolic activities in Arabidopsis cell culture. Planta 227, 57-66.

Raterink, R.J., Lindenburg, P.W., Vreeken, R.J., and Hankemeier, T. (2013). Three-Phase Electroextraction: A New (Online) Sample Purification and Enrichment Method for Bioanalysis. Analytical chemistry.

Rochfort, S.J., Trenerry, V.C., Imsic, M., Panozzo, J., and Jones, R. (2008). Class targeted metabolomics: ESI ion trap screening methods for glucosinolates based on MSn fragmentation. Phytochemistry 69, 1671-1679.

Rojas-Cherto, M., Peironcely, J.E., Kasper, P.T., van der Hooft, J.J., de Vos, R.C., Vreeken, R., Hankemeier, T., and Reijmers, T. (2012). Metabolite identification using automated comparison of high-resolution multistage mass spectral trees. Analytical chemistry 84, 5524-5534.

Savitski, M.M., Sweetman, G., Askenazi, M., Marto, J.A., Lang, M., Zinn, N., and Bantscheff, M. (2011). Delayed fragmentation and optimized isolation width settings for improvement of protein identification and accuracy of isobaric mass tag quantification on Orbitrap-type mass spectrometers. Analytical chemistry 83, 8959-8967.

Wichitnithad, W., McManus, T.J., and Callery, P.S. (2010). Identification of isobaric product ions in electrospray ionization mass spectra of fentanyl using multistage mass spectrometry and deuterium labeling. Rapid Commun Mass Spectrom 24, 2547-2553.

Wishart, D.S., Knox, C., Guo, A.C., Eisner, R., Young, N., Gautam, B., Hau, D.D., Psychogios, N., Dong, E., Bouatra, S., et al. (2009). HMDB: a knowledgebase for the human metabolome. Nucleic Acids Res 37, D603-610.