Unambiguous Natural Product ID

Unambiguous molecular formula identification of natural products by analyzing Isotopic Fine Structures acquired from high resolution mass spectrometer

Introduction

Identification of an exact molecular formula for natural products represents one of the major goals and challenges in discovery of novel secondary metabolites.[1] Molecular formulas can be utilized as a golden standard for dereplication of known natural products from existing compound databases, which prevents unnecessary time and reduces labor cost. In addition, the rapid confirmation of molecular formula would provide significant information for structural elucidation of unknown compounds and accelerate the whole discovery process of natural products.

Keywords: Mass Spectrometry, High Resolution, Isotopic Fine Structure, Molecular Formula Identification, Natural Products

Authors: Jia-Xuan Yan; Dr. Fan Zhang, Dr. Navid Adnani, Dr. Tim S. Bugni
School of Pharmacy, University of Wisconsin, Madison, USA

Mass spectrometry is a well-established technique for measuring the accurate mass to charge *(m/z)* ratios of different ions. This technique has been widely used in determining the molecular formulas of natural products as well as synthetic compounds. In mass spectrometry, resolution could be defined as: [2]

Resolution = R = M/ΔM

M is the *m/z* ratio of the selected peak and Δ*M* is usually defined as the peak width at its half-maximum peak height. The resolution of the measurement could be improved by increasing the resolving power of the mass spectrometers. Thus, the monoisotopic mass peak could be identified correctly and lead to an accurate molecular weight of the compound. However, the accurate mass alone is not sufficient for determining the unambiguous molecular formula of natural products due to combinatorial explosion.[3] It has been demonstrated that even with 0.1 ppm mass accuracy for the MS instrument, a unique molecular formula could not be determined when C, H, N, S, O, P atoms were included in the search list for molecules with molecular weight above 185.9760 Da.[3] Therefore, additional information, such as isotope abundance ratio, would be required for the determination.

Table 1: Mass defects and natural abundance of common isotopes in organic compounds

Element	Isotope	Atomic Mass (u)	Mass Defect (u)	Natural abundance (%)
Hydrogen	^{1}H	1.00783	0.00783	99.9885
	^{2}H (D)	2.01410	0.01410	0.0115
Carbon	^{12}C	12.00000	0.00000	98.93
	^{13}C	13.00335	0.00335	1.07
Nitrogen	^{14}N	14.00307	0.00307	99.632
	^{15}N	15.00011	0.00011	0.368
Oxygen	^{16}O	15.99491	-0.00509	99.757
	^{17}O	16.99913	-0.00087	0.038
	^{18}O	17.99916	-0.00084	0.205
Sulfur	^{32}S	31.97207	-0.02793	94.93
	^{33}S	32.97146	-0.02854	0.76
	^{34}S	33.96787	-0.03213	4.29
Chlorine	^{35}Cl	34.96885	-0.03115	75.78
	^{37}Cl	36.96590	-0.03410	24.22
Bromine	^{79}Br	78.91834	-0.08166	50.69
	^{81}Br	80.91629	-0.08371	49.31

The concept of mass defect is the result of different nuclear binding energies of different elements and their nuclides.[4] Conventionally, ^{12}C was defined as the element with zero mass defect while other nuclides have different mass defect depending on their relative nuclear binding energy to ^{12}C. Additionally, each element would have a certain ratio of different isotopes based on their natural abundance. Therefore, compounds with different elemental composition would have different exact masses as well as unique Isotopic Fine Structure (IFS). Mass defects and natural abundance of common elements in organic compounds and their isotopes are displayed in Table 1.[4]

Here we demonstrate the unambiguous molecular formula determination of a natural product echinomycin A using the above mentioned IFS concept. This method would potentially be used as a powerful tool for rapid discovery of novel compounds and quick dereplication of known compounds in natural product libraries. We recently found that IFS was crucial for determining an exact molecular formula for the unambiguous molecular formula determination of keyicin, an antibiotic produced by marine bacterial co-culture.[5] The determination of the exact formula played a crucial role in the structural elucidation process of keyicin because an exact carbon count by NMR was unattainable without ^{13}C isotopic enrichment.

Experimental

The examined compound, echinomycin, was isolated from the marine *Streptomyces sp.* WMMC-592 following the typical include the natural product isolation and purification process.[6] HPLC purification was performed using a Shimadzu LC-20AP system with a Luna 5 µm C18 100Å 250*10 mm column. Linear gradient from 25:75 MeCN/H_2O (with 0.1% acetic acid) to 50:50 MeCN/H_2O (with 0.1% acetic acid) over 35 minutes was used.

Mass spectrometry detection was performed using a Bruker 12T MRMS (Magnetic Resonance Mass Spectrometry) instrument and a Bruker QTOF MS instrument. The instruments were operated under ESI positive mode to acquire full scan MS spectra. Bruker ESI-MS tuning mix was used for the instrument calibrations. The compound was dissolved in LC/MS grade MeOH (2 µg/mL). The acquired data was analyzed by Bruker Compass DataAnalysis 4.4 SR1. Molecular formula determination was carried out using SmartFormula and the isotopic patterns were simulated by Simulate Pattern.

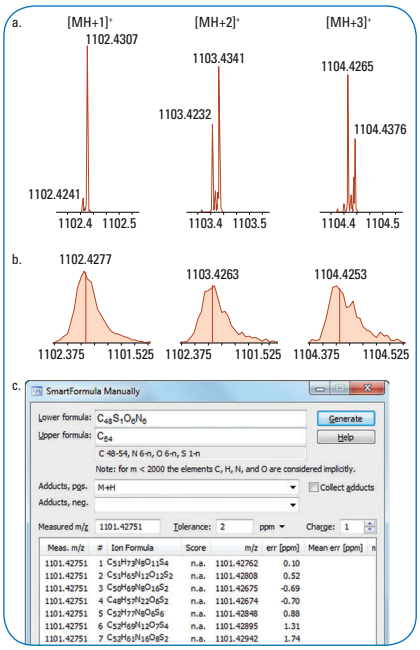

Figure 1: a.) Bruker 12T MRMS spectra and b.) Bruker QTOF MS spectra of tested compounds. [MH+1]+, [MH+2]+, [MH+3]+ ions were displayed. c.) SmartFormula analysis of possible adducts formulas.

Results and Discussion

Assignment of the molecular ion and SmartFormula prediction of possible molecular formulas

Direct infusion MRMS data collected under ESI (electrospray ionization) positive mode displayed an intense $[M+H]^+$ adduct peak with *m/z* ratio of 1101.4275. Isotopologues were fully resolved in the spectra for $[MH+1]^+$, $[MH+2]^+$ and $[MH+3]^+$ ions (Figure 1a). On the other hand, data acquired on a QTOF did not show resolved isotopologues (Figure 1b). SmartFormula was used to provide a range of reasonable formulas. The search was set to be C48S1N6O6 and the upper formula C54 based on preliminary NMR data. The MS error tolerance was set to be 2 ppm, and seven possible molecular formulas were given by SmartFormula analysis (Figure 1c): $[C_{51}H_{73}N_8O_{11}S_4]^+$, $[C_{51}H_{65}N_{12}O_{12}S_2]^+$, $[C_{50}H_{69}N_8O_{16}S_2]^+$, $[C_{48}H_{57}N_{22}O_6S_2]^+$, $[C_{52}H_{77}N_8O_6S_6]^+$, $[C_{52}H_{69}N_{12}O_7S_4]^+$ and $[C_{51}H_{61}N_{16}O_8S_2]^+$.

Determination of the exact molecular formula

Different element composition of the molecules would show their unique IFS and these fine structures could be observed with the improved resolution of the MRMS instrument (Figure 1a). For $[MH+1]^+$ ion, the difference between the two peaks was 0.0066 u, which matched the mass defect difference of $^{15}N^{12}C$ and $^{14}N^{13}C$ (0.0063 u). Similar calculations were performed and the different peaks observed in $[MH+2]^+$ and $[MH+3]^+$ ions were assigned to the combinations of $^{34}S^{12}C_2/^{32}S^{13}C_2$ and $^{34}S^{13}C^{12}C_2/^{32}S^{13}C_3$ respectively. In contrast, the spectra collected from the QTOF instrument were not able to display similar fine structures (Fig. 1b). Therefore, the accurate molecular formula could be identified from various possibilities by matching its unique IFS to the collected spectra. The IFS of three possible species, $[C_{51}H_{73}N_8O_{11}S_4]^+$, $[C_{48}H_{57}N_{22}O_6S_2]^+$ and $[C_{51}H_{65}N_{12}O_{12}S_2]^+$ were simulated by Simulate Pattern function and the simulated MS peaks were overlaid with the actual spectra (Figure 2). For $[MH+2]^+$ ion, the relative abundance of $[^{12}C_{51}H_{73}N_8O_{11}{}^{32}S_3{}^{34}S]^+$ was significantly higher than the actual molecule, (Figure 2a) which indicated the tested compound should have less sulfur atoms than $[C_{51}H_{73}N_8O_{11}S_4]^+$. Similar trend was found in the $[MH+3]^+$ ion overlaid spectra. The tested compound was suggested to contain 2 sulfur atoms since the sulfur related peaks in $[MH+2]^+$ and $[MH+3]^+$ overlaid spectra matched the simulated IFS of $[C_{48}H_{57}N_{22}O_6S_2]^+$ (Figure 2b). However, the simulated $[C_{48}H_{57}{}^{14}N_{21}{}^{15}NO_6S_2]^+$ ion abundance and $[^{12}C_{47}{}^{13}CH_{57}{}^{14}N_{21}{}^{15}NO_6S_2]^+$ ion abundance were higher than the corresponding ion abundance in the actual spectra, indicating this adduct formula was a mismatch. Similar evaluations were performed on $[C_{51}H_{65}N_{12}O_2S_2]^+$ and the simulated fine structure could match the actual spectra (Figure 2c), indicating the molecular formula of the tested compound was $C_{51}H_{64}N_{12}O_2S_2$. This compound was identified as echinomycin A by detailed NMR analysis.[7]

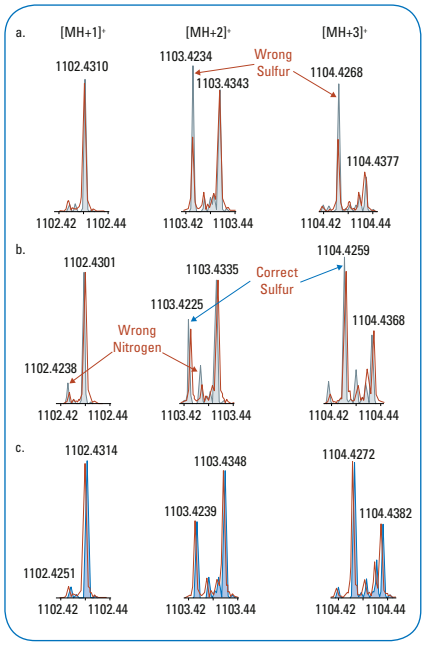

Figure 2: Overlaid IFS of the actual tested compound spectra and a.) $[C_{51}H_{73}N_8O_{11}S_4]^+$; b.) $[C_{48}H_{57}N_{22}O_6S_2]^+$; c.) $[C_{51}H_{65}N_{12}O_2S_2]^+$;

Conclusion

The unambiguous identification of the tested compound molecular formula among various possibilities was achieved by acquiring data from Bruker 12T MRMS system and analyzing the IFS of the acquired MS spectra. This technique features efficient and rapid molecular formula identification for further dereplication and novel natural products discovery efforts.

The SmartFormula and Simulate Pattern functions of Bruker Compass DataAnalysis 4.4 SR1 software were crucial for this analysis. When setting up proper limitations in SmartFormula search, the possible molecular formulas of the tested sample could be narrowed down and simplify the Simulate Pattern process. The identified accurate molecular formula would potentially be used as one of the key elements for metabolites database analyses.

Learn More

You are looking for further Information? Check out the Link or scan the QR Code.

www.bruker.com/isotopic-fine-structure-webinar

References

[1] Dührkop, K.; Hufsky, F.; Böcker, S., Mass Spectrom (Tokyo) 2014, 3(3): S0037; DOI: 10.5702/massspectrometry.S0037

[2] Guan, S.; Marshall, A. G., Anal. Chem. 1996, 68(1), 46-71.

[3] Kind, T.; Fiehn, O. BMC Bioinformatics 2006, 7:234.

[4] Sleno, L., J. Mass. Spectrom. 2012, 47, 226-236.

[5] Adnani, N.; Chevrette, M. G.; Adibhatla, S. N.; Zhang, F.; Yu, Q.; Braun, D. R.; Nelson, J.; Simpkins, S. W.; McDonald, B. R.; Myers, C. L.; Piotrowski, J. S.; Thompson, C. J.; Currie, C. R.; Li, L.; Rajski, S. R.; Bugni, T. S. ACS Chem. Biol. 2017, 12, 3093-3102.

[6] Zhang, F.; Adnani, N.; Vazquez-Rivera, E.; Braun, D.R.; Tonelli, M.; Andes, D.R.; Bugni, T.S. J. Org. Chem. 2015, 80, 8713-8719.

[7] Dell, A.; Williams, D. H.; Morris, H. R.; Smith, G. A.; Feeney, J.; Roberts, G. C., J. Am. Chem. Soc. 1975, 97(9), 2497-2502.

Bruker Daltonics GmbH & Co. KG

Bremen · Germany
Phone +49 (0)421-2205-0

Bruker Scientific LLC

Billerica, MA · USA
Phone +1 (978) 663-3660

ms.sales.bdal@bruker.com - www.bruker.com