Bruker Daltonics

Technical Note # TN-23

Certainty in Small Molecule Identification by Applying SmartFormula 3D on a UHR-TOF Mass Spectrometer

Break through the limitations of formula elucidation with new dimensions of accurate mass data

Introduction

Characterization of compounds is daily work in all laboratories working with products from synthetic chemistry, combinatorial libraries or targeted structural modifications in medicinal chemistry. Electrospray ionization time-of-flight (ESI-TOF) mass spectrometry has become an integral part of this process to determine accurate masses of the compounds. Finding the correct molecular formula of an unknown compound requires more information than mass accuracy alone. The number of possible molecular formulae increases exponentially with increasing m/z values, even when using small mass error tolerances of <2 ppm, defining the minimum carbon content, and using normal chemical rules for C/H ratios. [1] The presence of additional elements, beyond C,H,N,O, further adds to the formulae list.

How SmartFormula works

The number of candidate formulae is greatly reduced if the isotopic pattern is used as an additional, orthogonal criterion. The theoretical isotopic pattern of each potential accurate mass formula is compared with the experimental determined pattern using the SmartFormula algorithm, which examines the isotopic ratios and spacing of the complete isotopic pattern. The result is a short list of consistent formulae ranked in order of the fit. True isotopic pattern information over a wide dynamic range with fast spectral acquisition, together with accurate mass, are absolute necessities for this concept to work reliably. Bruker ESI-TOF mass spectrometers are superbly designed to meet these requirements.

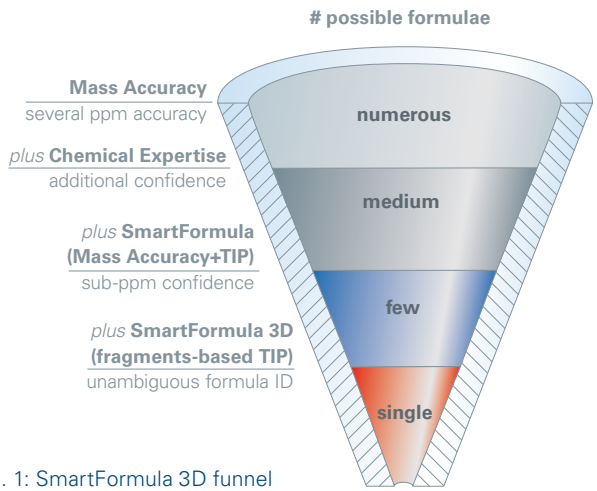

Fig. 1: SmartFormula 3D funnel

Real world example using mass accuracy only

Meas. m/z	#	Formula	m/z	err [ppm]	rdb	N-Rule	e⁻ Conf	mSigma
734.4683	1	C 34 H 60 N 11 O 7	734.4672	-1.6	10.5	ok	even	6.51
734.4683	2	C 37 H 68 N O 13	734.4685	0.3	4.5	ok	even	8.19
734.4683	3	C 35 H 56 N 15 O 3	734.4685	0.2	15.5	ok	even	18.68
734.4683	4	C 39 H 189 N 2 O 3	734.4693	1.3	-53.5	ok	even	23.88
734.4683	5	C 24 H 185 N 14 O 4	734.4698	2.0	-60.5	ok	even	47.49
734.4683	6	C 26 H 197 O 14	734.4698	2.0	-71.5	ok	even	49.24
734.4683	7	C 23 H 189 N 10 O 8	734.4684	0.1	-65.5	ok	even	52.47
734.4683	8	C 20 H 52 N 27 O 4	734.4690	0.9	8.5	ok	even	53.31
734.4683	9	C 22 H 64 N 13 O 14	734.4690	0.9	-2.5	ok	even	54.81
734.4683	10	C 22 H 193 N 6 O 12	734.4671	-1.7	-70.5	ok	even	61.52
734.4683	11	C 20 H 181 N 20 O 2	734.4671	-1.7	-59.5	ok	even	61.79
734.4683	12	C 21 H 68 N 9 O 18	734.4677	-0.9	-7.5	ok	even	65.90
734.4683	13	C 19 H 56 N 23 O 8	734.4677	-0.9	3.5	ok	even	67.10
734.4683	14	C 50 H 60 N 3 O 2	734.4680	-0.4	22.5	ok	even	82.59
734.4683	15	C 10 H 197 N 8 O 19	734.4690	0.8	-83.5	ok	even	120.69
734.4683	16	C 5 H 48 N 39 O 5	734.4695	1.6	1.5	ok	even	122.78
734.4683	17	C 9 H 72 N 11 O 25	734.4695	1.6	-20.5	ok	even	125.51
734.4683	18	C 5 H 177 N 32 O 3	734.4676	-1.0	-66.5	ok	even	130.04
734.4683	19	C 8 H 185 N 22 O 9	734.4689	0.8	-72.5	ok	even	130.75
734.4683	20	C 9 H 201 N 4 O 23	734.4676	-1.0	-88.5	ok	even	131.79
734.4683	21	C 7 H 60 N 25 O 15	734.4695	1.6	-9.5	ok	even	132.89
734.4683	22	C 4 H 52 N 35 O 9	734.4682	-0.2	-3.5	ok	even	135.80
734.4683	23	C 8 H 76 N 7 O 29	734.4682	-0.2	-25.5	ok	even	137.02
734.4683	24	C 7 H 189 N 18 O 13	734.4676	-1.0	-77.5	ok	even	140.74
734.4683	25	C 6 H 64 N 21 O 19	734.4682	-0.2	-14.5	ok	even	145.83

Spectrum Data | Mass List | Deconvolution Results | SmartFormula

Fig. 2: A 2 ppm window without additional filtering gives a list of 25 candidates.

Four candidates left using SmartFormula

#	Meas. m/z	Formula	m/z	err [ppm]	N-Rule
1	734.4683	C 34 H 60 N 11 O 7	734.4672	-1.6	ok
2	734.4683	C 37 H 68 N O 13	734.4685	0.3	ok
3	734.4683	C 35 H 56 N 15 O 3	734.4685	0.2	ok
4	734.4683	C 50 H 60 N 3 O 2	734.4680	-0.4	ok

trum Data

Fig. 3: After applying chemical rules 4 hits remain in a 2 ppm tolerance window.

Now, an extra dimension of information: SmartFormula 3D

For larger molecules, or those with many elements, a unique formula may not be possible within the mass spectral dimension. Adding MS/MS fragment information brings a whole new dimension of information into play (Fig. 1). Candidate formulae for fragment ions are used as small subunits of the parent ion. SmartFormula 3D iterates through the possible formulae that could be formed from these subunit candidates. As with SmartFormula, user input includes additional elements (or C,H,N,O by default), optional upper and lower limits for elements, number of rings and double bonds, C/H ratios, and whether to allow odd-electron fragment ions. Many potential formulae derived from the parent ion alone are eliminated, to leave only the candidate formulae that fit to the SmartFormula results for both the parent and matched fragment ions. The formulae for these ions, together with neutral losses, are displayed for review. SmartFormula 3D considers either all detected fragment ions, or only user selected ions, and candidates may be filtered to show only those with required neutral losses.

High mass formula determination

For compounds at higher mass, many hundreds of formulae are consistent with a mass accuracy of even 1 ppm or less. Even with prior knowledge of the compound, so that elemental formulae limits and chemistry rules may be applied, there are still many possible candidates. A confident determination of a compound such as erythromycin at m/z 734 (Fig. 2) is not possible. In fact, the SmartFormula pattern fitting gives the correct formula as the second best fit of 4 candidates within 2 ppm, using normal chemical rules (Fig. 3, 4). The fragment ion spectrum of erythromycin (Fig. 5) has 2 prominent fragment ions. The SmartFormula 3D interpretation of the spectrum considers the masses and isotopic patterns for all the fragment ions. The candidate formula list is reduced to only one possible formula; the correct formula for erythromycin ($C_{37}H_{68}NO_{13}$) (Fig 6). In addition, SmartFormula 3D shows that for the incorrect candidate formulae, prominent fragment ions in the spectrum do not match to that formula.
This unique formula assignment is only possible with accurate mass of < 2 ppm in both MS and MS/MS, True Isotopic Pattern (TIP™), and a low neutral loss mass tolerance. If the tolerances are broadened, then other candidate formulae are still possible.

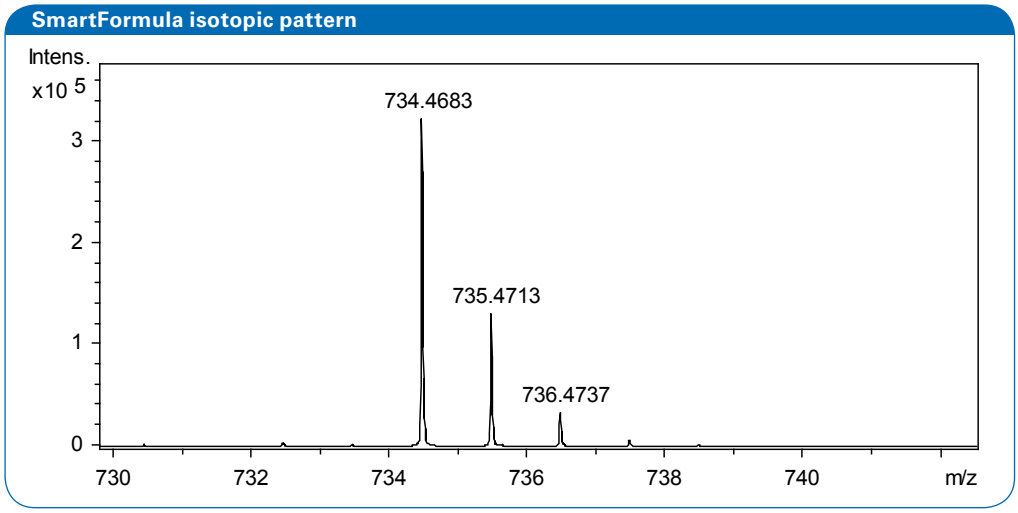

Fig. 4: Measured isotopic pattern of erythromycin.

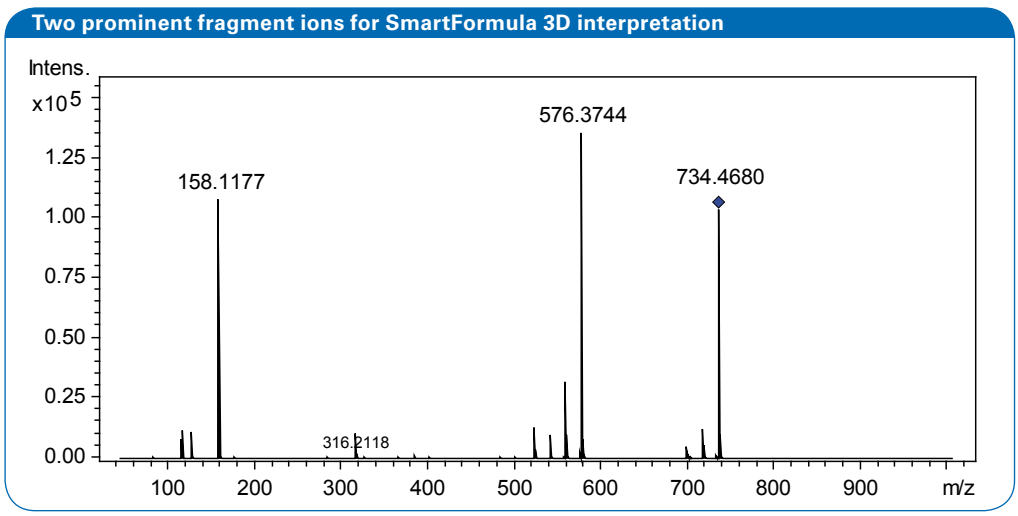

Fig.5: MS/MS spectrum of erythromycin.

Confirmation of unexpected impurities

In a different experiment, metabolites and impurities of alprenolol ($C_{15}H_{23}NO_2$) were detected in an LC/MS analysis, including an unexpected compound with m/z 188.0930 [2]. Only 1 candidate, $C_9H_{10}N_5$, was possible from SmartFormula, but chemically unlikely. To confirm the formula, MS/MS analysis was used. SmartFormula 3D confirms that the fragment ions at m/z 104.04 and 146.07 fit only to the formulae C_7H_6N and $C_8H_8N_3$, so the parent ion formula is correct (Fig. 7). This unexpected formula was searched against a chemical database; the most probable compound is benzoguanamine, used in thermosetting resins, which is an unexpected contaminant in this sample.

Conclusion

SmartFormula and SmartFormula 3D are standard, integral modules of the Bruker Daltonics DataAnalysis data processing software, part of the Compass 1.3 suite. SmartFormula automates formula generation for all the detected peaks in an LC/MS analysis and results are included in reports. SmartFormula 3D is intended for investigation and interpretation of individual MS/MS spectra, ideal for the user who is investigating structural elucidation challenges. With these advanced capabilities, instrument accuracy plays a less dominant role in formula determination than the isotopic pattern. However, for large and complex molecules, a small mass tolerance helps to reduce possibilities. Typical mass errors are < 2 ppm for our micrOTOF series products and < 1 ppm for the maXis.

SmartFormula 3D defines the correct formula

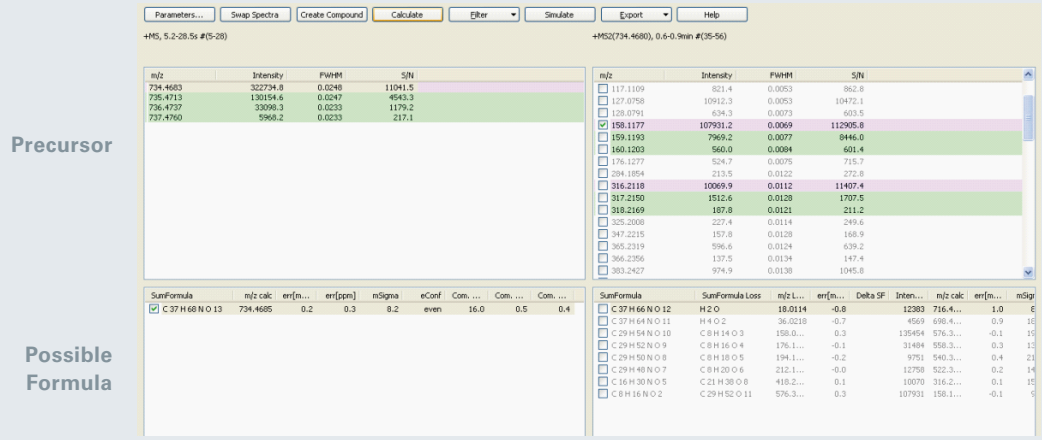

Fig. 6: Only the correct sum formula of erythromycin remains after SmartFormula 3D processing.

Detection of impurity after chemical synthesis

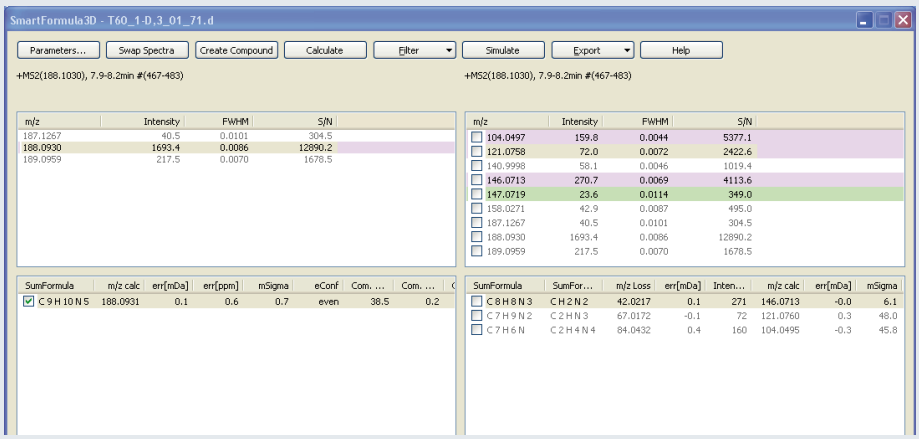

Fig. 7: Only the correct sum formula of benzoguanamine remains after SmartFormula 3D processing.

Acknowledgement

SmartFormula 3D was developed from an original concept by Don Richards, Pfizer.

References

[1] Kind, Fiehn: Metabolomic database annotations via query of elemental composition: Mass accuracy is insufficient even at less than 1 ppm, BMC Bioinformatics 2006, 7:234.

[2] Personal communication with Don Richards, Pfizer UK

Authors

Catherine Stacey, Sebastian Goetz, Carsten Baessmann, Bruker Daltonics.

Keywords	Instrumentation & Software
ab initio formula determination	SmartFormula 3D
Small Molecule Identification	maXis

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

www.bdal.com

Bruker Daltonics Inc.

Billerica, MA · USA
Phone +1 (978) 663-3660
Fax +1 (978) 667-5993
ms-sales@bdal.com

Bruker Daltonik GmbH

Bremen · Germany
Phone +49 (421) 2205-0
Fax +49 (421) 2205-103
sales@bdal.de