Using the Agilent LC/MSD TOF to Identify Unknown Compounds

Application Note

Doug McIntyre
Agilent Technologies

Introduction

In the pharmaceutical industry, it is not unusual to generate small quantities of unknown compounds in addition to the intended products when synthesizing lead molecules, when doing larger-scale synthesis of promising candidates, or during manufacturing. While HPLC with a UV detector can detect the presence of unknowns, mass spectrometry is usually required to positively identify them. Since the synthesis or manufacturing process generally provides some clues to the composition of unknowns, it is frequently possible to propose and/or confirm a logical structure. One approach to identifying these unknowns is to interpret the combination of precursor ion spectra and product ion spectra produced by MS/MS analysis. An alternate approach is to use the accurate mass capabilities of a time-of-flight (TOF) mass spectrometer. Accurate mass TOF systems are used to confirm target compounds,[1] but can also be used to identify unknowns.

While a quadrupole mass spectrometer is typically operated with unit mass resolution and can assign the mass to the nearest 0.1 *m/z*, modern TOF systems produce spectra with a resolution of 4000 to 10,000, depending on the mass of the ion, and can assign a mass to better than 5 ppm accuracy. This has significant implications when trying to propose possible empirical formulas. Table 1 shows what happens when you consider a molecule such as reserpine (MW 608.2734) and restrict the elemental composition to combinations of C, H, O, and N.

Table 1. Number of theoretical formulas for a compound composed of C, H, O, and N with a molecular weight 608.2734

Mass Accuracy (ppm)	Possible Formulas
165 (quadrupole)	209
10	13
5	7
3	4
2	2

Note that even with 2 ppm mass measurement accuracy, there are still two possible molecular formulas. Furthermore, a unique formula does not translate to a unique structure. Typically, other information such as knowledge of the synthesis, other spectral data, or product ion spectra from collision-induced dissociation (CID) is needed for unambiguous identification.

This note describes the application of the Agilent LC/MSD TOF for the identification of unknowns in synthetic mixtures.

Experimental

The system used was an Agilent LC/MSD TOF coupled to an Agilent 1100 Series HPLC. The TOF was equipped with a dual-nebulizer electrospray source. Combined with the automated calibrant delivery system (CDS) built into the LC/MSD TOF, the dual-nebulizer source allows continuous introduction of a reference mass compound at a low concentration.

The instrument was scanned from *m/z* 100 to 1000 for most samples. This range included the two reference mass compounds, which produced ions at *m/z* 121.050873 and 922.009798. The instrument was scanned at one scan per second; with each scan composed of approximately 10,000 transients. The majority of samples were run at a fragmentor setting of 150 volts. This generally resulted in a strong signal for the molecular ions with minimal fragmentation. Some experiments were done alternating the fragmentor between 150 volts and 215 volts. The higher setting produced ions resulting from in-source CID. In these cases, the number of transients per scan was reduced to 5000 so that a scan rate of one scan cycle per second was still maintained.

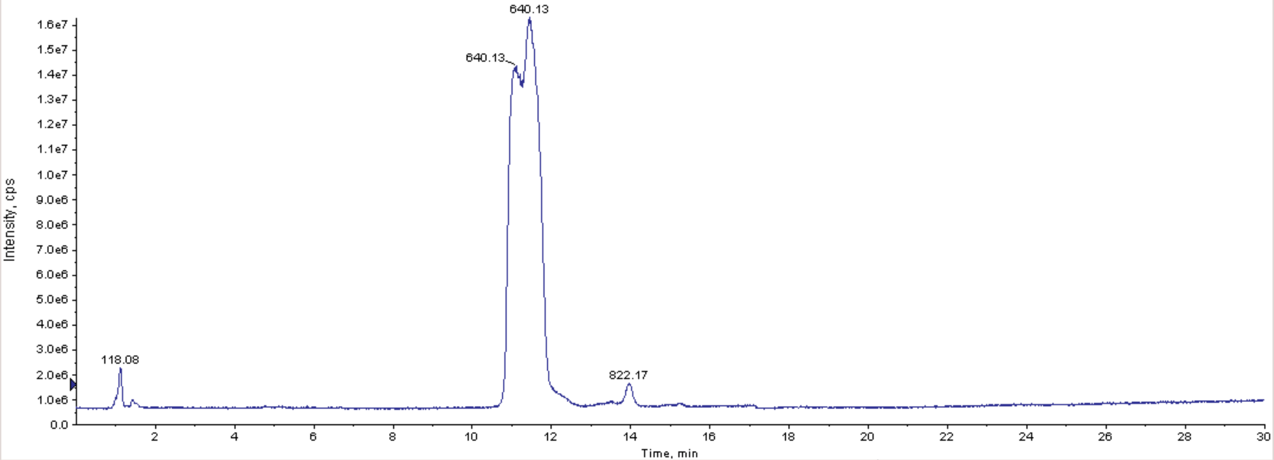
The data system stored data from a diode array detector as well as the mass spectral data, allowing for several approaches to identifying the presence of unknowns. For the UV data, either the total wavelength chromatogram (TWC) was used or a signal was extracted for a specific wavelength range.

For the MS data, the LC/MSD TOF software used the signals from the reference mass ions to automatically correct the mass assignments in each spectrum prior to the spectrum being stored to disk. Thus, the spectra were already reference mass corrected when viewed in the data analysis software.

Total ion chromatograms (TICs), base peak chromatograms (BPCs), and extracted ion chromatograms (EICs) were all used. BPCs and EICs covering a mass range narrower than the complete scan range often gave signals that more clearly showed the presence of trace compounds. Figure 1 shows an example of a TIC (a) and an EIC (b) from the same data. The vertical scale on the EIC is also expanded 10 times.

Spectra from the appropriate chromatogram were averaged and suspected molecular ions were selected. Using the elemental composition calculator built into the data analysis software, and applying constraints on the elemental composition, possible formulas were determined for simulated unknowns, where the actual compound was known, and for true unknowns that were trace components in synthetic mixtures.

a. Total Ion Chromatogram



b. Extracted Ion Chromatogram, *m/z* 500–900 (expanded 10X vertically)

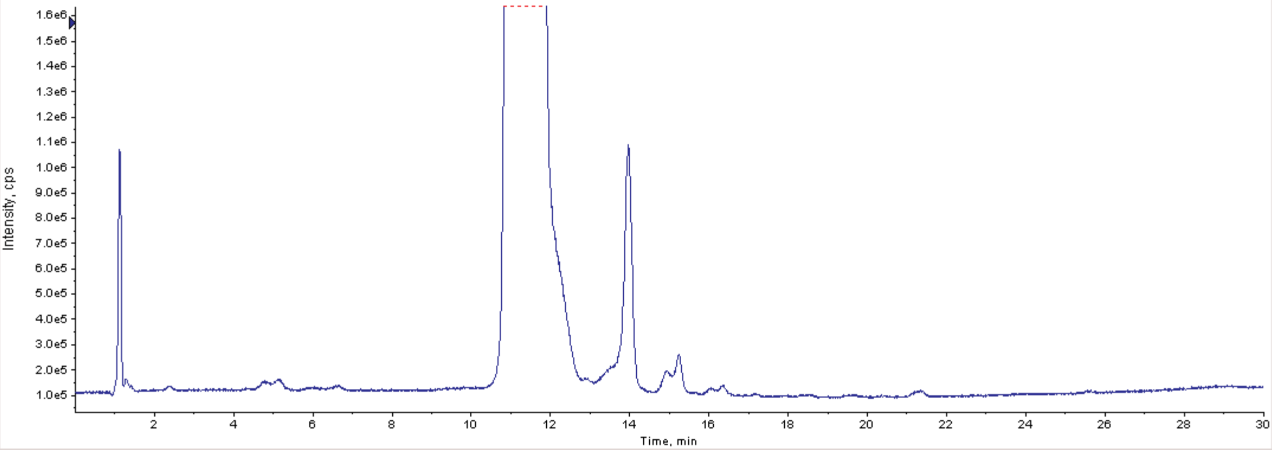


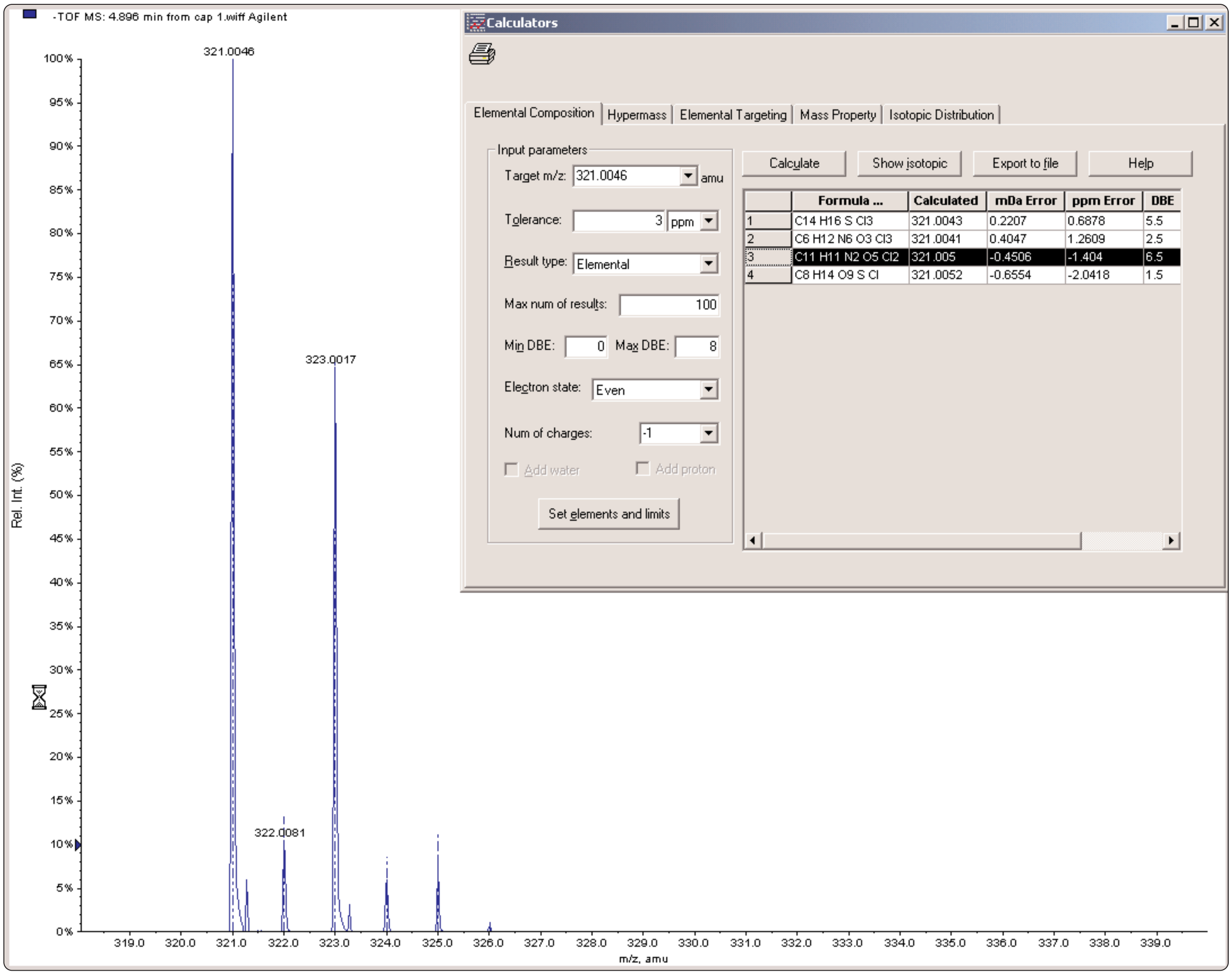
Figure 1. TIC and EIC of a mixture with lower-level unknown components

The Agilent LC/MSD TOF is specified to provide mass accuracy to better than 3 ppm for the compound reserpine. This accuracy was used as the tolerance setting in the elemental composition calculator. In addition, the electron state was set to Even, since these were API spectra and API ionization almost always produce ions with even numbers of electrons. A correct electron state setting is critical since ignoring the mass of an electron represents a 1 ppm error for a compound with a molecular weight around 500 u. The number of charges was set to +1 or –1 depending on the ionization polarity being used. The possible elements and their limits were set based on prior knowledge of the mixture being analyzed.

Results and Discussion

The first sample analyzed was the antibiotic, chloramphenicol, which was treated as an unknown in order to demonstrate the general technique. The elements were restricted to C, H, O, N, S and Cl and the number of charges was −1 because the compound was analyzed in negative ion mode.

The elemental composition calculator produced four possible formulas based on the ion at *m/z* 321.0046 as shown in Figure 2. In order to narrow the choices, the isotope ion at *m/z* 322.0081 should be consistent with the number of carbons in the molecule and the naturally occurring abundance of the ^{13}C isotope. In addition, the ion at *m/z* 323.0017 should reflect the naturally occurring abundances of ^{34}S and ^{37}Cl. Selecting “Show Isotopic” from the calculator will actually overlay the theoretical isotope abundances on the spectrum. Taking these criteria into account, the first formula, $C_{11}H_{11}N_2O_5Cl_2$ turned out to be the correct formula.



Input parameters

Target m/z: 321.0046 amu

Tolerance: 3 ppm

Result type: Elemental

Max num of results: 100

Min DBE: 0 Max DBE: 8

Electron state: Even

Num of charges: -1

☐ Add water ☐ Add proton

	Formula ...	Calculated	mDa Error	ppm Error	DBE
1	C14 H16 S Cl3	321.0043	0.2207	0.6878	5.5
2	C6 H12 N6 O3 Cl3	321.0041	0.4047	1.2609	2.5
3	C11 H11 N2 O5 Cl2	321.005	-0.4506	-1.404	6.5
4	C8 H14 O9 S Cl	321.0052	-0.6554	-2.0418	1.5

Figure 2. Possible formulas generated from chloramphenicol data by the elemental composition calculator

In addition to isotope information, adduct information can be used to reduce the number of possible formulas. For example, a mixture of four sulfa drugs: sulfamethizole, sulfamethazine, sulfachloropyridazine, and sulfadimethoxine, gave a chromatographic peak that produced a mass spectrum with ions at *m/z* 311.0814 and 333.0624 as shown in Table 2. Using the elemental composition calculator the following possible formulas were proposed, assuming the *m/z* 333.0624 ion could contain sodium.

Table 2. Possible ion formulas consistent with observed ions

Ion at *m/z* 311.0814		**Ion at *m/z* 333.0624**	
A	$C_5H_{15}N_{10}O_2S_2$	D	$C_{10}H_9N_{10}O_2S$
B	$C_{12}H_{15}N_4O_4S$	E	$C_{12}H_{14}N_4O_4SNa$
C	$C_{13}H_{11}N_8S$		

Choice A has too few carbons to match the isotope data. Choices B and E are the only formulas that are consistent with a protonated and sodiated adduct.

Even in instruments that are not capable of MS/MS, in-source CID can yield structural information that can be used to aid in the identification of unknowns. An analysis of the sulfonamide antibiotic, sulfachloropyridazine, demonstrates the usefulness of this approach. The LC/MSD TOF can acquire alternating scans at different fragmentor voltages. By using values of 150 and 215 volts, spectra with and without fragmentation were obtained across the chromatographic peak representing sulfachloropyridazine. Figure 3 shows a resulting mass spectrum. In addition to the protonated and sodiated ions at *m/z* 285.0207 and 307.0027, there is a fragment ion at *m/z* 156.0016. By observing the isotope ratios, it is possible to

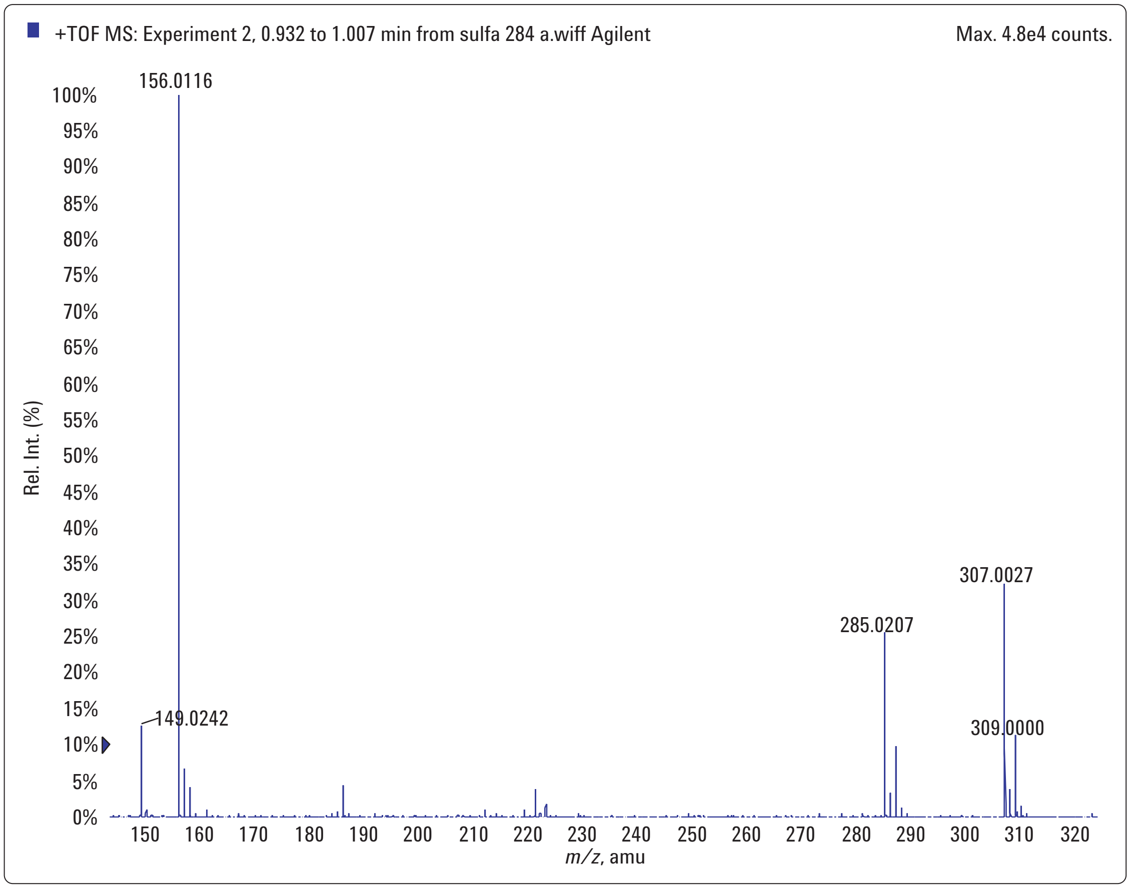


Figure 3. Sulfachloropyridazine spectrum with adduct and fragment ions

narrow down the number of proposed formulas from seven to one. Since the ion at *m/z* 156.0016 must come from the parent ion, its list of possible elements is based on the five elements (C, H, N, O, S) that make up sulfachloropyridazine. With this constraint, the calculator only comes up with a single formula, $[C_6H_6NO_2S]^+$ to match the data.

Conclusions

The identification of unknown components in synthetic mixtures is made much easier by the Agilent LC/MSD TOF. Mass accuracy of better than 3 ppm results in a fewer potential formulas for most compounds. When other information such as isotope ratios, fragment ions, and adduct ion are considered, it is often possible to propose a unique empirical formula.

References

1. “Automated Empirical Formula Confirmation Using the Agilent LC/MSD TOF,” Agilent Technologies Application Note, 2003, publication number 5989-0625EN.

www.agilent.com/chem

Printed in the U.S.A. February 11, 2004
5989-0626EN

Agilent Technologies