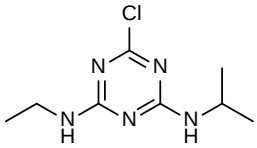
Application of spectral accuracy to improve the identification of organic compounds in environmental analysis

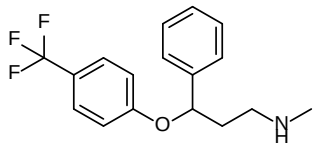
Emmanuel Eysseric[1], Killian Barry[1], Francis Beaudry[2], Magali Houde[3], Christian Gagnon[3], Pedro A. Segura[1,*]

* Tel: 1-(819) 821-7922. Fax: 1-(819) 821-8019. E-mail: pa.segura@usherbrooke.ca
[1] Department of Chemistry, Université de Sherbrooke, Sherbrooke, QC J1K 2R1
[2] Groupe de Recherche en Pharmacologie Animal du Québec (GREPAQ), Department of Veterinary Biomedicine, Université de Montréal, Saint-Hyacinthe, QC J2S 2M2
[3] Environment and Climate Change Canada, Montreal, QC H2Y 2E7

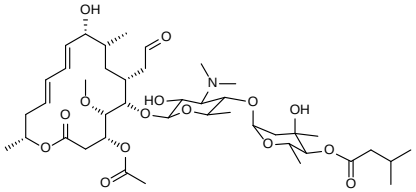
Supporting Information



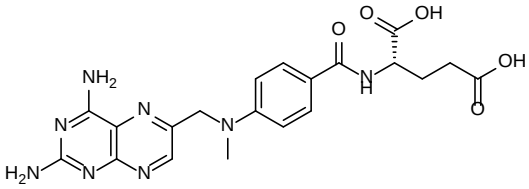
Atrazine (ATZ)
$C_8H_{14}N_5Cl$ (215 Da)



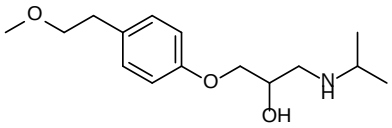
Fluoxetine (FLX)
$C_{17}H_{18}F_3NO$ (309 Da)



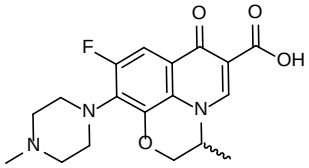
Josamycin (JOS)
$C_{42}H_{69}NO_{15}$ (827 Da)



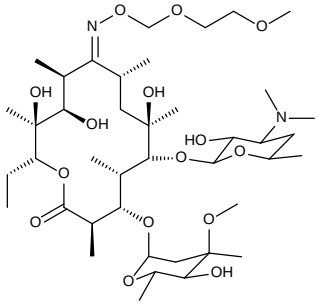
Methotrexate (MTX)
$C_{20}H_{22}N_8O_5$ (454 Da)



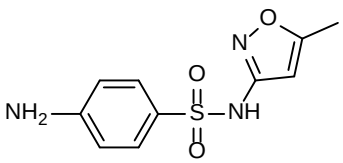
Metoprolol (MET)
$C_{15}H_{25}NO_3$ (267 Da)



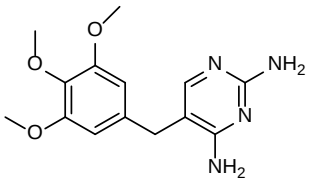
Ofloxacin (OFL)
$C_{18}H_{20}N_3O_4F$ (361 Da)



Roxithromycin (ROX)
$C_{41}H_{76}N_2O_{15}$ (837 Da)



Sulfamethoxazole (SMX)
$C_{10}H_{11}N_3O_3S$ (253 Da)



Trimethoprim (TRI)
$C_{14}H_{18}N_4O_3$ (290 Da)

Figure SI-1. Acronym, neutral nominal mass and molecular structure of the test compounds used in this study.

SI-1. External calibration and stability of the QqQMS system

A sodium formate solution was used for external calibration and four of the test compounds (MTP, FLX, OFL and JOS) were used as internal calibrants for ATZ, SMX, TRI, MTX and ROX. Figure SI-2 shows the impact of the calibration method. As expected, mass accuracy significantly improved when internal calibration was applied. In the case of ATZ, it improved from 524 ppm (113 mDa) to 6 ppm (1.3 mDa), allowing to get a better ranking (from 3 ± 1 to 1st) even with a relatively low spectral accuracy (in both cases ≤ 91%).

Mass accuracy stability of the QqQMS using external calibration was evaluated by successive injections of a mixture of ATZ, SMX and TRI spiked at 300 μg L^{-1} in MeOH during a 12 h period (Figure SI-3). The results showed that the average mass accuracy was 74 ± 17 mDa for ATZ, 66 ± 30 mDa for SMX and 46 ± 23 mDa for TRI. Those results indicated that the QqQ mass spectrometer was relatively stable overtime when using external calibration, however the average mass accuracy was too high, between 77 and 113 mDa, which negatively affects the ranking of the correct formula. It was not possible to obtain results for ROX using external calibration because its large mass error and large molecular mass lead to a high number of generated formulas. Major improvement of the results as illustrated by ATZ (Figure SI-2) were observed for all compounds using internal calibration.

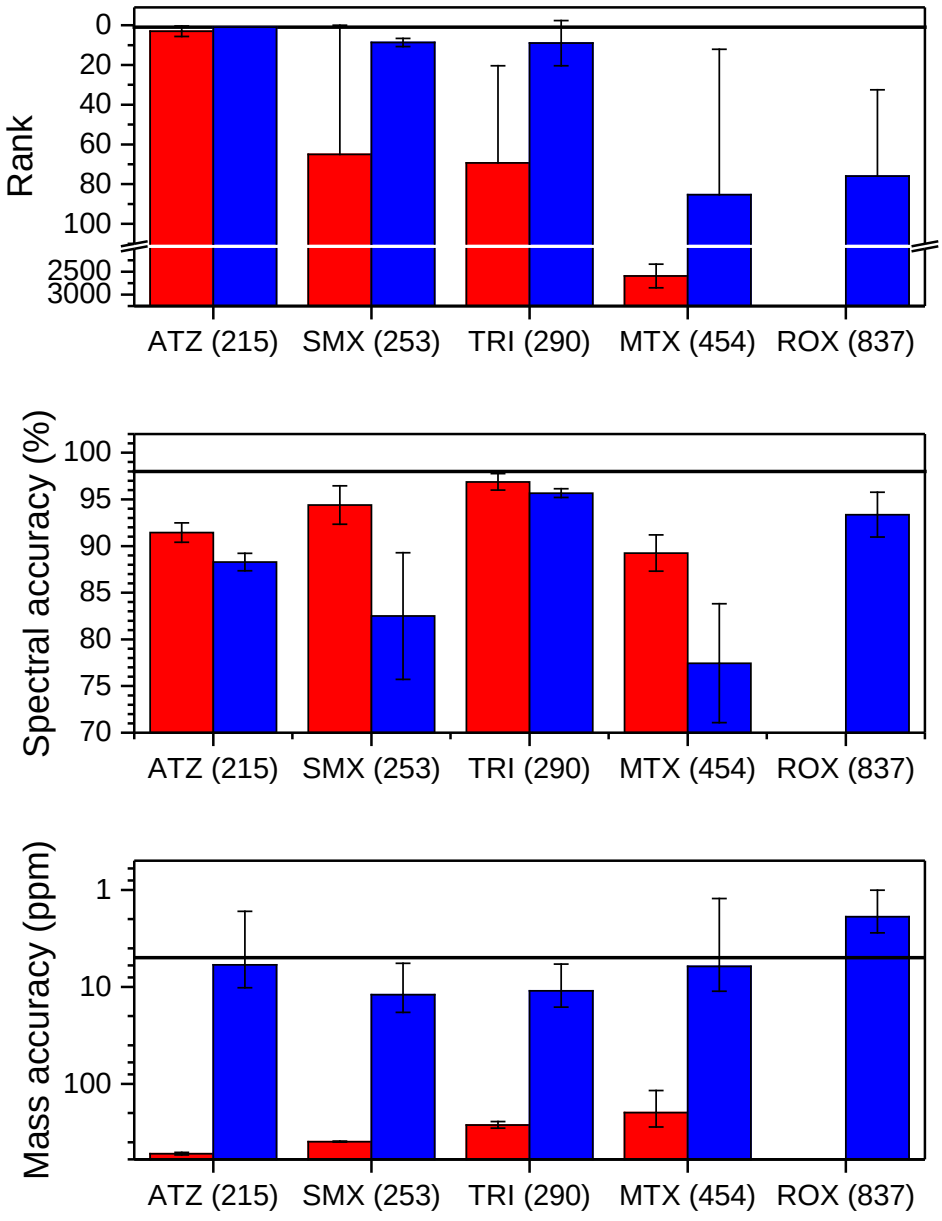


Figure SI-2. Ranking (top), spectral accuracy (middle) and mass accuracy (bottom) of the test compounds at a concentration of 300 μg L^{-1} in MeOH measured with the QqQMS using external (red) and internal calibration (blue). Straight lines indicate: expected value for ranking, 1st (top); the threshold of high spectral accuracy, 98% (middle) and accepted value for maximum mass accuracy, 5 ppm (bottom).

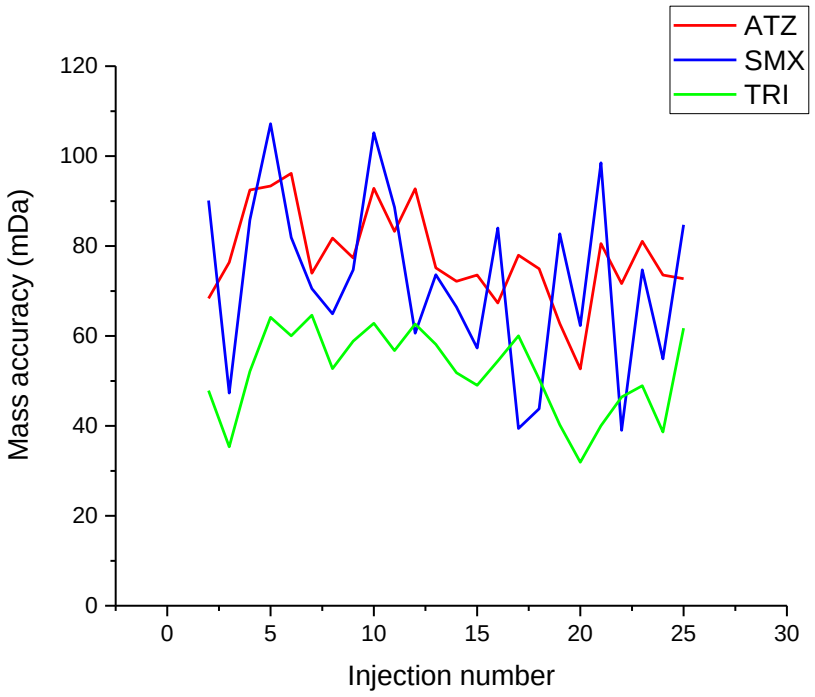


Figure SI-3. Stability of the mass accuracy obtained with external calibration in the QqQMS instrument over a 12 h period for ATZ, SMX and TRI spiked at 300 µg L^{-1} in MeOH.

A previous study demonstrated that post-acquisition analysis with MassWorks of low resolution data, obtained with gas chromatography-quadrupole mass spectrometry using perfluorotributylamine as external calibrant, can be employed to identify unknown compounds [1]. Therefore, a second experiment of stability of mass accuracy over time was performed by injecting a solution of sodium formate each hour within a period of 12 h in the QqQMS. In those experiments the solution of sodium formate at 0.5 µM was used in chromatographic conditions to study the evolution of mass accuracy in the QqQMS over time. Chromatographic conditions were the following: mobile phase flow rate was 50 µLmin^{-1}, and the mobile phase was a mixture of 2-propanol and water (9:1, v/v). Run time was 1 hour and injection volume was 10 µL. No chromatographic column was used as there was no need for separation. A union was used instead of the column. QqQMS source and ion optics parameters were the same as those used for the analysis of the selected compounds. In total 11 injections were made over a period of 12 h. Sodium formate spectra were then calibrated post-acquisition with MassWorks CLIPS algorithm. The second injection spectrum was used to calibrate the rest of the data.

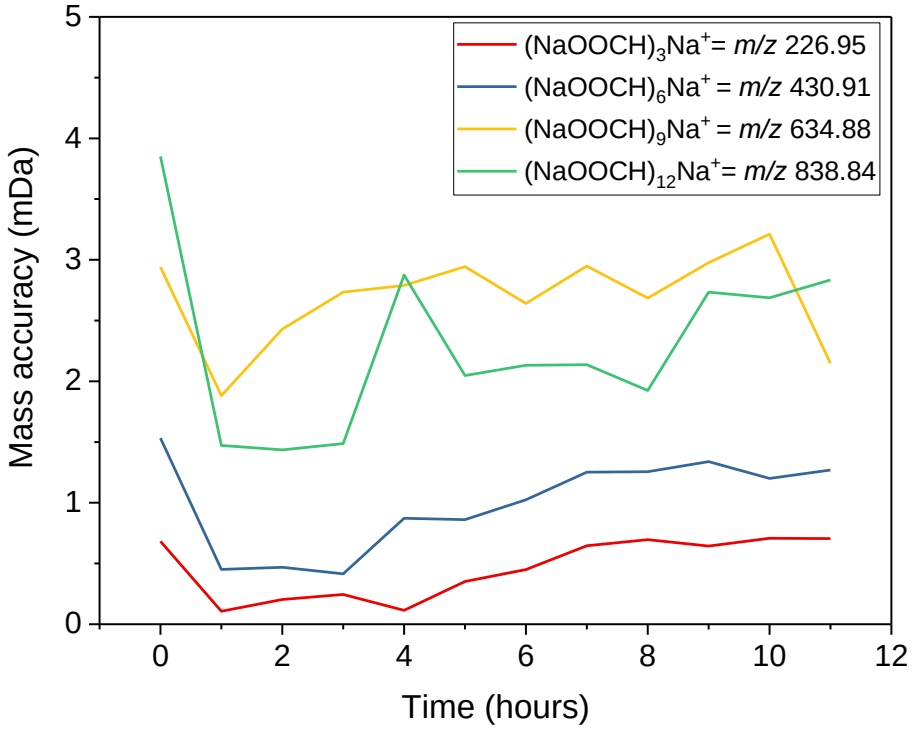


Figure SI-4. Post-acquisition CLIPS-calibrated stability of the mass accuracy obtained with external calibration in the QqQMS instrument over a 12 h period for sodium formate at 0.5 μM in 2-propanol- water (9:1, v/v). Injection #2 (time = 1 hour) was used to calibrate all other masses.

As can be seen in Figure SI-4, post-acquisition CLIPS-calibrated masses showed little variation (difference between maximum and minimum values was <3 mDa) and had good mass accuracy ($\Delta m \leq 4$ mDa) over the span of 11 hours. This is a large discrepancy compared to the external mass calibration results for ATZ, SMX and TRI discussed previously (Figure SI-3). It could be partly explained by completely different chromatographic conditions: mobile phase composition, flow rate, and separation (or the absence thereof) were all dissimilar. It is also worth noting that all masses were calibrated with their own *m/z* values contrary to the external calibration in which the *m/z* values of the calibrants were close but different to those of the test compounds. Hence, at this point is not clear why a major systematic error in the mass accuracy of the QqQMS data with external calibration could not be corrected by the software. That issue is out of the scope of the present work and will be the topic of future research.

SI-2. Error in spectral accuracy determination

For a reported spectral of certain value, the spectral error is (100 – spectral accuracy). The square of the spectral error follows a Chi-square distribution. Then the confidence interval (CI) of the spectral error is given by:

$$(100 - spectral\ accuracy) \times \sqrt{\frac{\chi^2_{\left(\frac{\alpha}{2}\right)}}{k-2}} < spectral\ error\ < (100 - spectral\ accuracy) \times \sqrt{\frac{\chi^2_{(1-\frac{\alpha}{2})}}{k-2}}$$

where χ^2 is the Chi-squared critical value (upper- or lower-trail) for a given significance level α and degrees of freedom and *k-2* represents the degrees of freedom that are calculated according to the number of data points in the profile isotopic pattern. A value of 2 is subtracted from the data points because of the baseline and the pure theoretical mass spectrum. For the spectral accuracy of JOS at 50 ng mL^{-1}, we would have spectral accuracy of 96.5686%, and 450 data points across the profile isotopic pattern, therefore the 95% confidence limits of the spectral error are:

$$Upper\ limit = (100 - 96.5686) \times \sqrt{\frac{536.6167}{448}} = 3.63\%$$

$$Lower\ limit = (100 - 96.5686) \times \sqrt{\frac{368.5509}{448}} = 3.21\%$$

SI-3. Interference rejection

It is expected that ions from coeluting matrix compounds could lower the spectral accuracy and the ranking if they are found in the spectral region of interest, i.e. between the peaks of the relevant isotope pattern. The interference rejection function in MassWorks was designed to correct those potential issues. This function allows the exclusion of a sub-spectral region from the calibrated experimental spectrum if the relative theoretical abundance in that region is less than a defined fractional value. For example, interference rejection with a value of 0.001 means that any spectral region where the relative theoretical abundance is less than 0.1%, relative to the most intense isotope, would be ignored and not factored into the spectral accuracy calculation. The effect of this function was measured and is displayed on Figure SI-5. At 80 $\mu g L^{-1}$ spiked in the matrix, the effect of interference rejection on the ranking was unnoticeable for most compounds except for MTX. However spectral accuracy improved slightly with interference rejection. Interference rejections results for the 300 $\mu g\ L^{-1}$ solution are presented in Figure SI-6 and Table SI-1 and followed a similar trend. Since the matrix can contain a myriad of compounds of low abundance, there are a multitude of peaks that can decrease the similarity between calibrated and theoretical isotopic patterns. Therefore, by rejecting those low abundance peaks that were not part of the isotopic pattern of the compound of interest, spectral accuracy increases. As expected, the effect of interference rejection was more important for larger compounds with more significant M+3 or M+4 peaks such as MTX, JOS and ROX. While the interference rejection is a useful tool to correct spectral accuracy calculations in MassWorks, it must be used with care and only when known interferences of *m/z* value close to the compound of interest co-elute. Otherwise use of this feature can lead to wrong conclusions in the identification process.

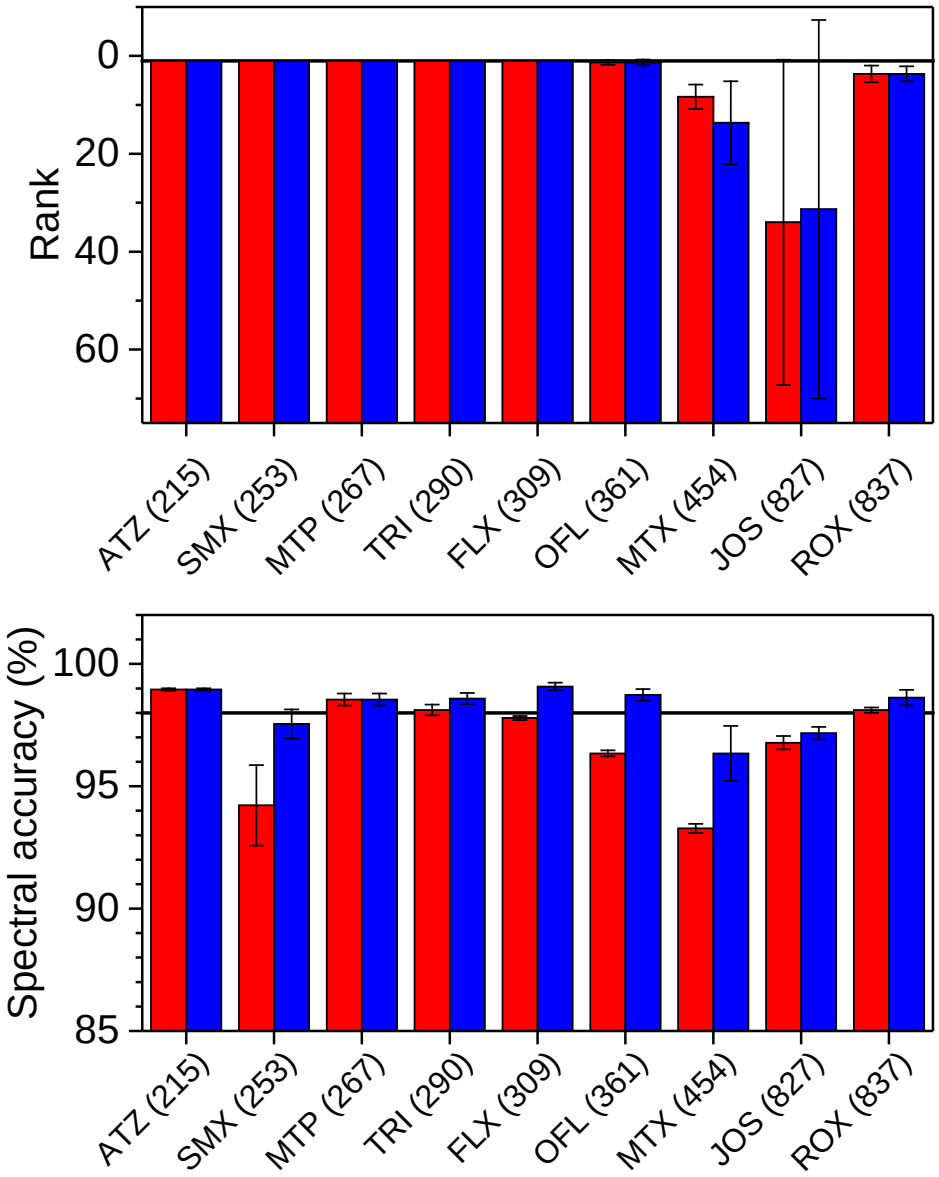


Figure SI-5. Rank and spectral accuracies measured with a QqTOFMS in the 80 μg L^{-1} matrix solutions with (blue) and without (red) interference rejection. Straight lines indicate: expected value for ranking, 1st (top) and threshold of high spectral accuracy, 98%(bottom).

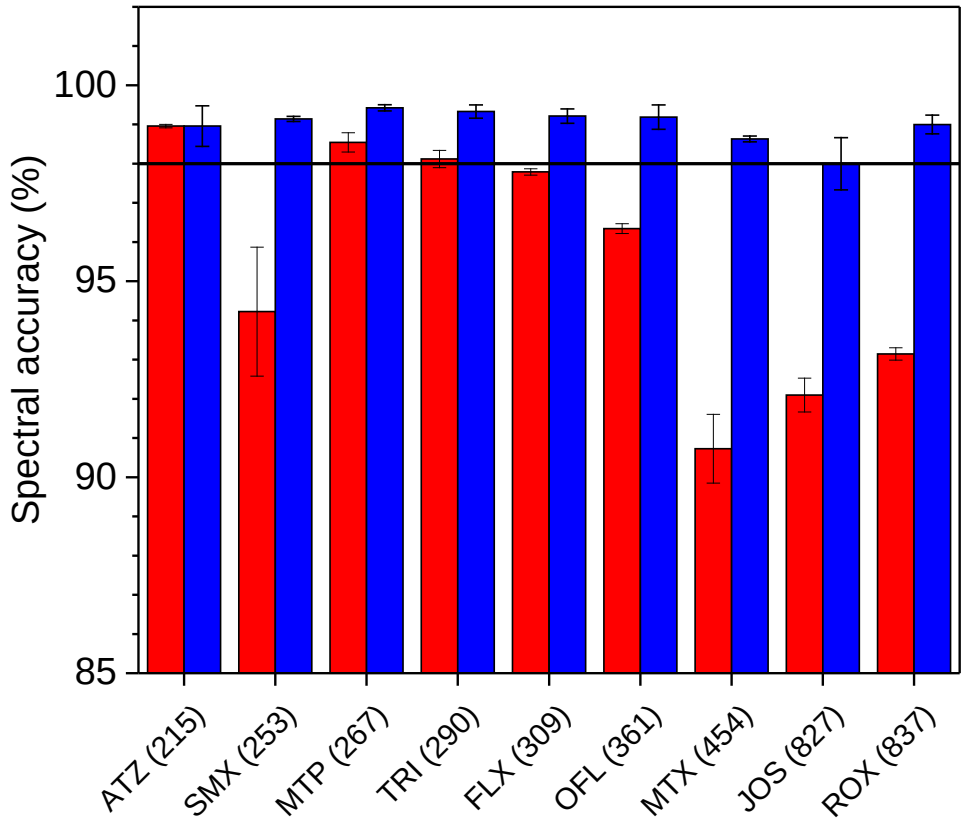


Figure SI-6. Spectral accuracies measured with a QqTOFMS in the 300 μg L^{-1} matrix solutions with (blue) and without (red) interference rejection. The straight line indicates the threshold of high spectral accuracy, 98%.

Table SI-1. Impact of interference rejection on correct formula ranking.

Compound	**Matrix spiked at 300 μg L^{-1}**	
	Without interference rejection	With interference rejection
ATZ (216)	1	1
SMX (254)	1	1
MTP (268)	1	1
TRI (291)	1	1
FLX (310)	1	1
OFL (362)	1	1
MTX (455)	2 ± 1	2 ± 1
JOS (827)	11 ± 12	13 ± 14
ROX (837)	2 ± 1	2 ± 1

Table SI-2. Determination of matrix effects in data acquired with the three mass spectrometers

Compound	Order of elution	Matrix/MeOH areas ratio 300 μg L^{-1} QqQMS	Matrix/ MeOH areas ratio 80 μg L^{-1} QqTOFMS (R_{FHWM}=25 K)	Matrix/ MeOH areas ratio 300 μg L^{-1} QqTOFMS (R_{FHWM}=25 K)	Matrix/ MeOH areas ratio 80 μg L^{-1} QqOrbitrapMS (R_{FHWM}=70 K)	Matrix/ MeOH areas ratio 300 μg L^{-1} QqOrbitrapMS (R_{FHWM}=70 K)	Matrix/ MeOH areas ratio 80 μg L^{-1} QqOrbitrapMS (R_{FHWM}=140 K)	Matrix/ MeOH areas ratio 300 μg L^{-1} QqOrbitrapMS (R_{FHWM}=140 K)
ATZ (215)	8	1.2 ± 0.1	0.832 ± 0.003	0.7 ± 0.1	0.96 ±0.03	1.04 ± 0.01	0.94 ± 0.03	1.04 ± 0.04
SMX (253)	5	2.8 ± 0.4	0.6 ± 0.4	0.49 ± 0.01	1.03 ± 0.02	1.02 ± 0.02	0.95 ± 0.02	0.99 ± 0.02
MTP (267)	4	NA	0.7 ± 0.1	0.78 ± 0.05	0.89 ± 0.01	1.02 ± 0.02	0.81 ± 0.02	1.03 ± 0.01
TRI (290)	2	1.6 ± 0.2	0.62 ± 0.02	0.72 ± 0.04	1.01 ± 0.04	1.09 ± 0.04	0.99 ± 0.03	1.02 ± 0.03
FLX (309)	6*	NA	0.9 ± 0.1	0.72 ± 0.03	0.93 ± 0.02	1.07 ± 0.02	0.87 ± 0.02	1.12 ± 0.05
OFL (361)	3	NA	2.8 ± 0.4	20 ± 2	4.9 ± 0.3	7.0 ± 0.2	6.1 ± 0.2	8.8 ± 0.3
MTX (454)	1	1.5 ± 0.2	3.6 ± 0.3	1.1 ± 0.2	1.04 ± 0.04	1.07 ± 0.02	1.04 ± 0.07	1.07 ± 0.04
JOS (827)	7	NA	0.6 ± 0.2	0.36 ± 0.02	0.83 ± 0.03	1.06 ± 0.03	0.79 ± 0.02	1.01 ± 0.02
ROX (837)	6*	0.90 ± 0.02	0.9 ± 0.1	0.72 ± 0.03	1.03 ± 0.04	1.05 ± 0.02	0.94 ± 0.04	1.07 ± 0.04

* These two compounds co-eluted. NA: Not available.

SI-4. Software comparison

SI-4.1 Methods

Molecular Formula Finder of the ChemCalc web application [2] (http://www.chemcalc.org/mf_finder) was used to determine the number of total formulas corresponding to specific accurate masses measured in the QqQMS. To perform a reasonable comparison, the parameters used were as similar as possible as those used for MassWorks. For example, allowed elements and their number were determined by MassWorks based on the seven golden rules and varied depending on the compound. Double bond equivalents (0 to 999) and reference values (2012) were the default values. Mass error (tolerance) was determined experimentally according to mass accuracy and values were between 1 and 4 mDa.

For QqTOFMS data, the built-in tool for Bruker's Data Analysis Smart Formula was used for formula determination. All parameters were equivalent for Smart Formula and MassWorks in order to have a meaningful comparison.

SI-4.2. Molecular formula finder and MassWorks using QqQMS data

One of the goals of using spectral accuracy is to reduce the number of potential formulas corresponding to an accurate mass within a given mass accuracy. Using data of the matrix spiked at 300 μg L^{-1}, the results obtained with MassWorks were compared to a tool that generates formulas only from mass accuracy, Molecular Formula Finder. As shown in Table S-3, results indicated that MassWorks allows a significant reduction (up to 96%) in the number of potential molecular formulas and lead to drastically improvement of the rankings. For example, accurate mass in the QqQMS of the protonated ion of TRI for the three replicates was *m/z* 291.1492, 291.1467 and 291.1454. Using those values and parameters indicated previously, Molecular Formula Finder listed 8380 ±45 possible formulas while MassWorks, based on both spectral accuracy and mass accuracy, only returned 336 ± 15.

It should be highlighted that MassWorks gives the ability to perform accurate mass measurements with a system that is not designed for this kind of experiments. Indeed, the QqQMS used is neither a high resolution nor a high-end mass analyzer with accurate mass capabilities. Therefore, the QqQMS instrument used would not be able to perform accurate mass measurements without the MassWorks software. Although the results were greatly improved using MassWorks, the obtained rankings were not good enough to allow formulae determination with a high degree of certitude. It is also important to keep in mind that these measurements were not possible with low concentration tested, 80 μg L^{-1}, spiked in the river extract. Therefore, signal intensity for the QqQMS data was critical. For example, MTX obtained a low ranking (260 ± 120) because signal had a low signal-to-noise ratio.

Table SI-3. Number of possible formulas and ranking of the test compounds in ChemCalc and MassWorks for the results obtained with the QqQMS using the matrix spiked at 300 μg L^{-1}.

Compound	Number of possible formulas		Rank	
	ChemCalc	MassWorks	ChemCalc	MassWorks
ATZ (215)	1662 ± 10	93 ± 6	230 ± 140	3 ± 2
SMX (253)	4763 ± 12	554 ± 21	360 ± 100	130 ± 100
TRI (290)	8380 ± 45	336 ± 15	>1000	9 ± 7
MTX (454)	142400 ± 1400	4978 ± 341	>1000	260 ± 120
ROX (837)	13800 ± 200	4485 ± 49	>1000	49 11

SI-4.3. Smart Formula and MassWorks using QqTOFMS data

Data acquired with the QqTOFMS using the test compounds spiked in the matrix were processed with SmartFormula, an algorithm developed by Bruker that uses a similar approach to MassWorks to determine the best molecular formula match. While details about the SmartFormula algorithm were not provided, it compares experimental and theoretical isotopic patterns without MS peak shape calibration by calculating a statistical match factor, the Sigma value [3]. Therefore, formulas associated to lower Sigma values are ranked higher since their theoretical isotopic patterns are more similar to the experimental isotopic pattern. As can be seen in Table S-4, both MassWorks and Smart Formula consistently ranked first the correct formula for test compounds < 350 Da (ATZ, FLX, MTP, SMX and TRI) at 80 μg L^{-1}. MassWorks showed significantly better formula ranking for compounds with molecular mass > 350 Da at 80 and 300 μg L^{-1} except for JOS at 300 μg L^{-1} (11 ± 12) which had a larger standard deviation than the ranking obtained with SmartFormula (11 ± 4). An inaccurate ranking value for a single injection was the cause of the high standard deviation observed for MassWorks. This latter value was due a significant discrepancy between the experimental and theoretical M+1 peaks, an effect lost in the ranking with Smart Formula. Based on these results, the performance of the formula determination algorithm of MassWorks was superior than the algorithm used by SmartFormula, since the former obtained better ranking of the correct formula of larger compounds (>350 Da) and it was more robust when using lower intensity signals.

Table SI-4. Accurate formula ranking with MassWorks and Smart Formula of the target compounds spiked in the river matrix.

Compound	Spiked at 80 μg L^{-1}		Spiked at 300 μg L^{-1}	
	MassWorks rank	Smart Formula rank	MassWorks rank	Smart Formula rank
ATZ (216)	1	1	1	1
SMX (254)	1	1	1	1
MTP (268)	1	1	1	1
TRI (291)	1	1	1	1
FLX (310)	1	1	1	1
OFL (362)	1 ± 1	4 ± 4	1	2 ± 1
MTX (455)	8 ± 2	38 ± 21	2 ± 1	8 ± 2
JOS (827)	34 ± 33	102 ± 73	11 ± 12	11 ± 4
ROX (837)	4 ± 2	34 ± 19	2 ± 1	19 ± 14

References

(1) Gu, B.; Wang, Y. *Spectroscopy* **2008**.
(2) Patiny, L.; Borel, A. *J. Chem. Inf. Model.* **2013**, *53*, 1223-1228.
(3) Bristow, T.; Constantine, J.; Harrison, M.; Cavoit, F. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 1213-1222.