Qualitative uncertainty (reliability) of chemical identification with High Resolution Mass Spectrometry

Marios Kostakis, Nikolaos S. Thomaidis

Laboratory of Analytical Chemistry, Department of Chemistry, National and Kapodistrian University of Athens, Panepistimioupolis Zografou, 157 71 Athens, Greece; emails – makostak@chem.uoa.gr; ntho@chem.uoa.gr

Abstract

High Resolution Mass Spectrometry (HRMS) is becoming more and more accessible and applied in routine analyses. The high resolving power and the excellent mass accuracy of HRMS make this technique an excellent choice for multi-analyte screening methods. Of course, HRMS is not lacking of errors and there are many examples of misidentification of compounds due to matrix effects, spurious errors, and inappropriate choice of screening parameters. Moreover, the identification criteria for modern HRMS have not been clearly documented yet. The estimation of uncertainty of identification (or **reliability**) is a way to assess the capabilities of identification of HRMS. There are two methodologies for the estimation of reliability of identification, the **contingency tables** and the **Bayesian methods**. In the first approach, reliability is estimated through the calculation of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) ratios. In the Bayesian approach, the reliability is estimated for the calculation of probabilities of false detect, but also considering historical and conditional probabilities. The aim of this study is to estimate the uncertainty of identification with both approaches and discuss the identification criteria of LC-QTOFMS using the uncertainty of identification, in order to minimize the false detects. Towards that aim, fish samples (sea bass and sea bream) were spiked with sulfonamides at different concentration levels, near to the limit of identification (LoI). The experiment was repeated in intermediate precision conditions and the uncertainty of identification was estimated from the results with both approaches. The identification criteria were evaluated and discussed.

Instrumentation

Column: Acclaim RSLC 120 C18
2.2 µm, 2.1 × 100 mm
Pre-column: VanGuard (Waters):
Acquity UPLC BEH C18 1.7 µm, 2.1 × 5 mm

Bruker, MaXis Impact
Ultra High Resolution
Time-of-Flight Mass Spectrometer

Identification Criteria

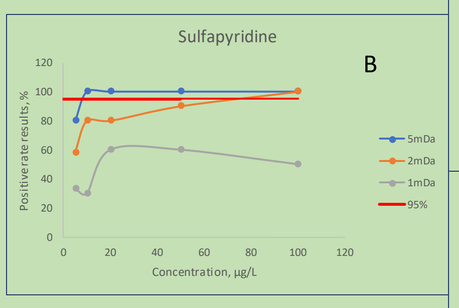
Mass Accuracy < 5mDa
Isotopic Fit score < 200 mSigma
Retention Time Tolerance (delta RT) < 0.2 min
Peak Area – Intensity levels (and ratio)
Fragment Ions

Experimental

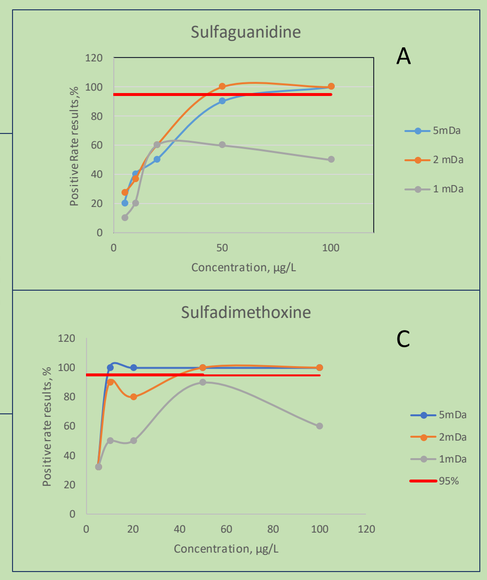
Matrix material: fish muscle
Spiked samples: 5 replicates x 2 days x 5 concentration levels
Concentration levels: 5, 10, 20, 50, 100 µg/kg

Identification Criteria

Mass Accuracy



Plot 1. Representative performance curve plots for 5 mDa, 2 mDa and 1 mDa mass tolerance.



Mass Accuracy was investigated in 5 mDa, 2 mDA and 1 mDA

The performance curves for every analyte were constructed (positive results rate *vs* concentration).

From performance curve was determined the Limit of Identification at 95%.

5 mDa: High number of positive results (high false positive rate)

1 mDa: Low number of positive results, high number of false negatives

Table 1. Percentage of positive result rates for all analytes

Level (µg/L)	5	10	20	50	100	mDa
Sulfaguanidine	0	10	**42**	100	100	2
sulfamethizole	10	27	82	**60**	100	2
Sulfachloropyridazine	**60**	100	100	100	100	5
Sulfaclozine	0	40	**44**	100	100	5
sulfadiazine	**90**	100	100	100	90	5
sulfamethoxazole	**90**	100	100	100	90	5
Sulfapyridine	58	80	80	**90**	100	2
sulfamerazine	63	90	**90**	100	100	5
sulfameter	0	0	40	**67**	100	2
sulfamethoxypyridazine	27	36	**60**	100	100	2
sulfamonomethoxine	0	20	18	**83**	100	2
sulfamoxole	80	70	71	**90**	100	2
sulfisoxazole	30	90	70	**90**	100	2
sulfadimidine	0	10	10	**80**	100	5
sulfadimethoxine	32	90	**80**	100	100	2
sulfadoxine	32	80	**90**	100	100	2
Sulfaquinoxaline	32	50	**80**	100	100	2

Isotopic Fit

- Isotopic fit score is a measure of the correlation between theoretical and measured isotopic pattern peak and expressed as mSigma value.
- Valid range: 0-1000.
- The lower is the mSigma , the better is the fitting
- Because of the limited number of atoms in organic compounds (C, O, N), it is necessary to use narrow window, but this increases the number of false negative results[1]
- Its mean value and deviation of every analyte in every concentration level was calculated

Table 2. Mean values and deviation of Isotopic fitting for every analyte at different concentration levels

mSigma	5 µg/kg		10 µg/kg		20 µg/kg		50 µg/kg		100 µg/kg	
	mean	deviation	mean	deviation	mean	deviation	mean	deviation	mean	deviation
Sulfaguanidine	613	30	591	55	586	90	512	62	409	139
sulfamethizole	123	46	75	28	76	32	45	13	14	6
Sulfachloropyridazine	205	133	172	166	119	57	75	77	46	57
Sulfaclozine	nm	nm	293	202	230	148	129	90	98	86
sulfadiazine	89	40	78	63	76	44	23	8	11	4
sulfamethoxazole	356	162	211	175	287	173	148	130	60	52
Sulfapyridine	114	13	54	29	45	24	9	5	6	3
sulfamerazine	463	125	264	112	271	141	67	29	42	16
sulfameter	nm	nm	67	34	48	13	26	11	10	3
sulfamethoxypyridazine	92	38	60	25	40	21	17	11	10	3
sulfamonomethoxine	nm	nm	69	29	44	20	25	10	10	4
sulfamoxole	315	142	165	69	186	162	42	75	8	4
sulfisoxazole	357	156	224	123	186	167	49	30	12	9
sulfadimidine	nm	nm	57	37	47	13	36	7	31	1
sulfadimethoxine	141	81	80	50	85	50	29	13	12	6
sulfadoxine	80	40	40	22	26	15	9	3	13	3
Sulfaquinoxaline	218	150	290	151	161	107	230	89	192	58

mSigma<200 | *300<mSigma<200* | *mSigma>300*

Retention Time

In all cases RT tolerance was lower than 0.1 min
Exceptions were sulfameter and sulfamethoxypyridazine. They are isomers with very close RT (difference of RT< 0.2 min) and the software confuses them or identify them twice.

Uncertainty

Table 3. Results of uncertainty for every analyte with both approaches.

	contigency table approach		Bayes approach	
	PPV	NPV	P(A|A)	P(nA|nA)
Sulfaguanidine	38.5	92.3	73.5	78.3
sulfamethizole	100	100	84.7	100
Sulfachloropyridazine	40.0	87.0	96.9	100
Sulfaclozine	80.0	80.0	69.4	75.9
sulfadiazine	97.5	10.0	99.6	14.3
sulfamethoxazole	96.0	0	100	0
Sulfapyridine	31.1	81.8	95.2	30.6
sulfamerazine	97.4	81.8	82.1	100
sulfameter	45.8	84.4	64.1	74.2
sulfamethoxypyridazine	62.5	100	81.4	100
sulfamonomethoxine	52.6	100	68.9	100
sulfamoxole	23.3	100.0	97.6	100
sulfisoxazole	36.8	100	92.6	100
sulfadimidine	100	80.4	35.9	91.4
sulfadimethoxine	48.8	92.3	73.5	78.3
sulfadoxine	47.6	100	96.7	100
Sulfaquinoxaline	57.1	92.3	73.5	78.3

Contingency Table approach
PPV: Positive Predictive Value, True positive results with respect to total positive results[2,3]
NPV: Negative Predictive Value, true negative results with respect to total negative results[2,3]
Bayes approach
P(A|A): Conditional probability of true positive results[2,3]
P(nA|nA):Conditional probability of true negative results[2,3]

Conclusions and Perspectives

- The identification criteria for HRMS, namely mass accuracy, Isotopic fitting score and retention time, were investigated.
- A mass accuracy of 2 mDa (and in some few cases, at 5 mDa) is the most appropriate value in order to avoid false detects.
- Isotopic fitting need a caution on identification, because it is dependent on analyte concentration and the elemental structure in order to be reliable.
- Retention time is a very reliable and stable criterion for identification
- The uncertainty for identification was calculated with both approaches (contingency tables and Bayes theory)
- As next step is the study of the mass fragmentation

Literature

1. A. Kaufmann, P. Butcher, K. Maden, S. Walker, M. Widmer, Analytica Chimica Acta, 856 (2015), 54-67.
2. L. Cuadros-Rodriguez et al, Trends in Analytical Chemistry, 80 (2016), 612-624.
3. A. Pulido, I. Ruisanchez, R. Boque, F.X. Rius, Trends in Analytical Chemistry, 22 (2003), 647-654.