Supporting Online Material

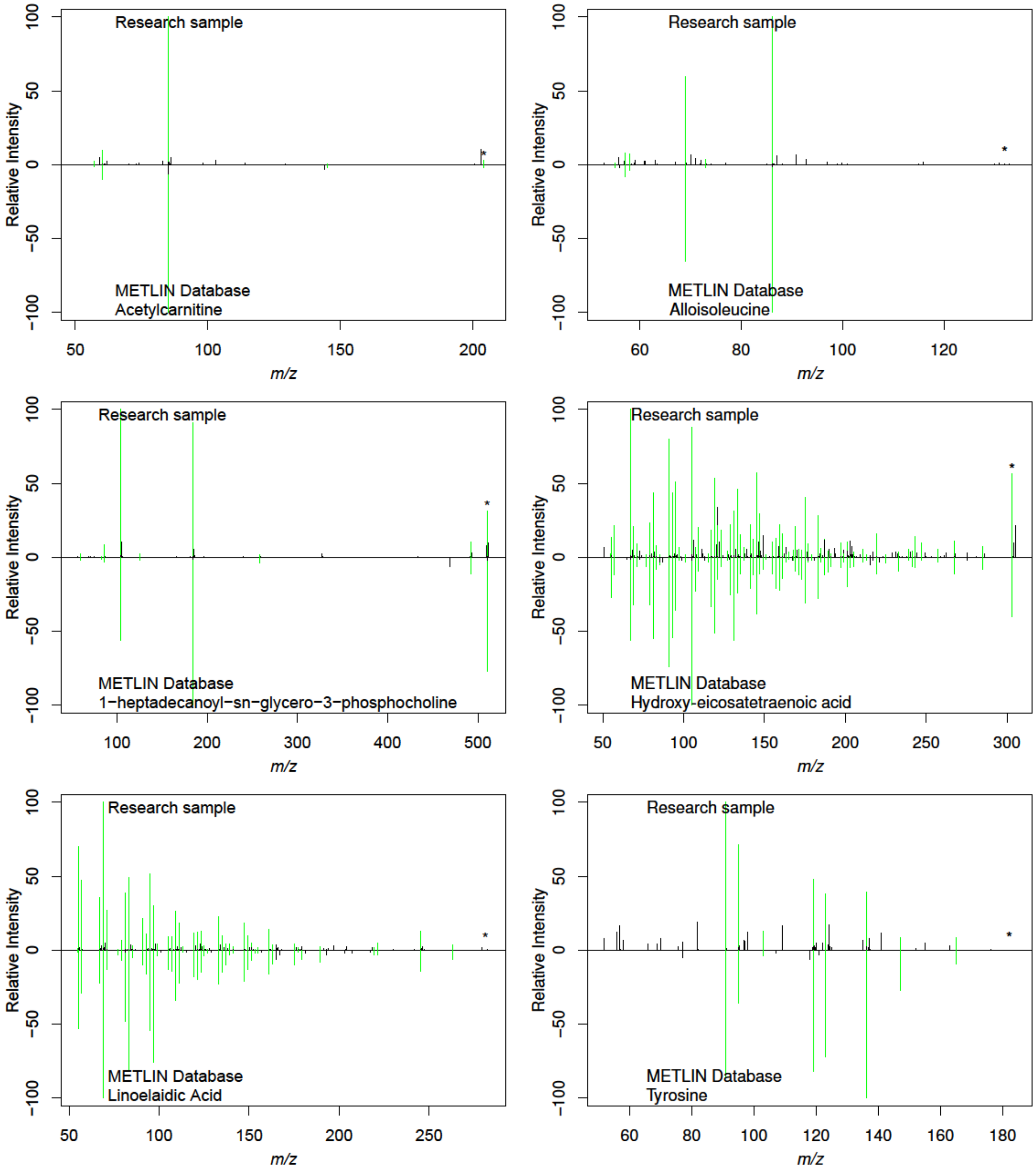
An accelerated workflow for untargeted metabolomics using the METLIN database

Ralf Tautenhahn, Kevin Cho, Winnie Uritboonthai, Zhengjiang Zhu, Gary J. Patti, and Gary Siuzdak

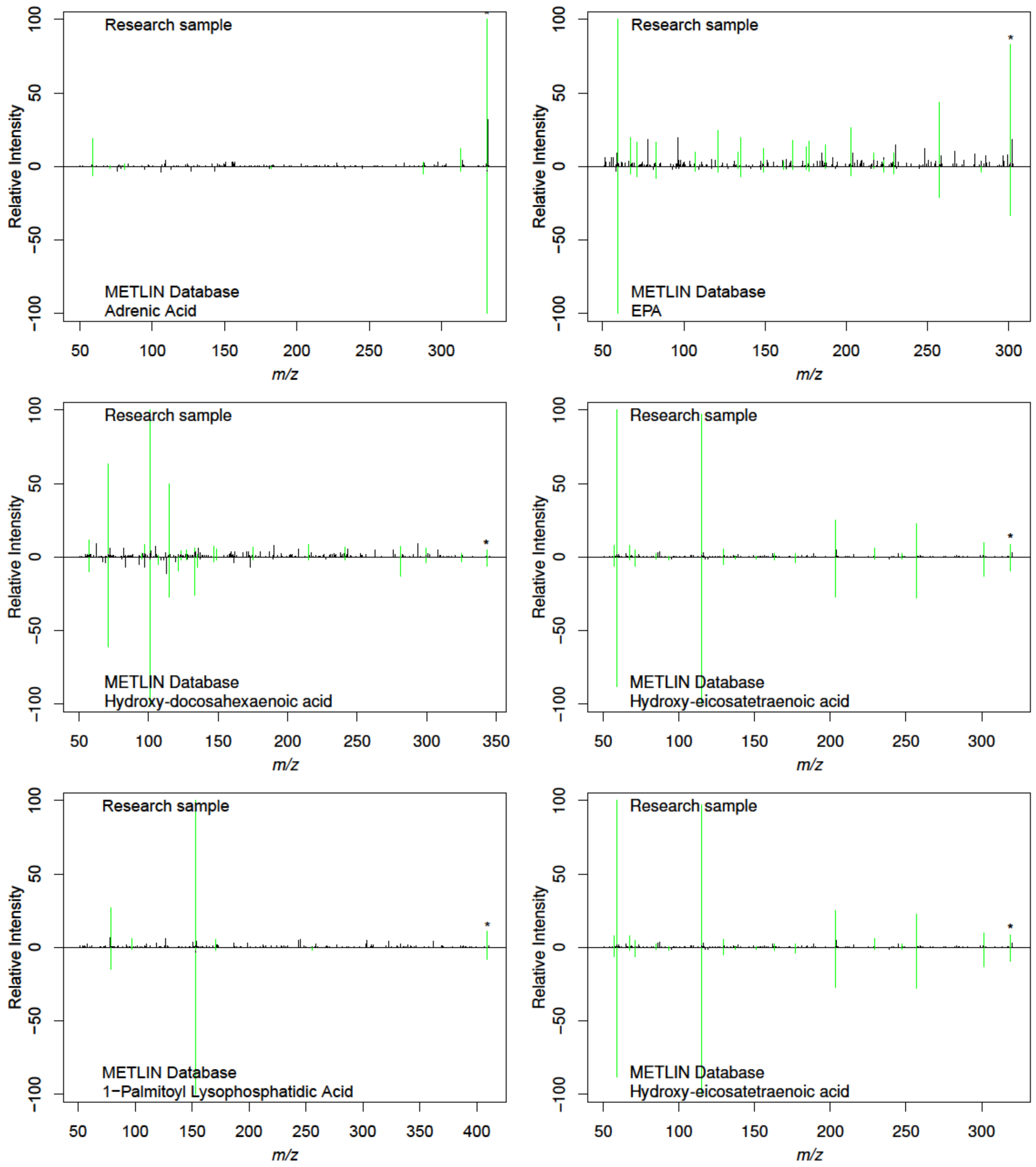
Materials and Methods

Human Serum from human male AB plasma, sterile-filtered (H4522) was purchased from Sigma. Metabolites were extracted from *E.coli* or serum by using cold methanol and acetone as described previously[1,2]. After sample collection, a volume of 600 μL of cold (-20 ºC) acetone was added to the each sample, vortexed for 30 s, and the sample incubated 1 min in liquid nitrogen. The samples were thawed at room temperature and incubated in liquid nitrogen two more times prior to a 10-min sonication. After 1 h at -20 ºC, the samples were centrifuged at 13,000 rpm for 15 min and the resulting supernatant was stored at -20 ºC. The precipitate was then mixed with 400 μL of methanol/water/formic acid in the ratio of 86.5/12.5/1.0 and sonicated prior to a 1-h incubation at -20 ºC. After centrifugation, the supernatant was collected and transferred to that which was collected after acetone extraction. The solution was dried with a vacuum concentrator (SpeedVac) at room temperature and redissolved in 100 μL of 95% acetonitrile/5% water for liquid chromatography/mass spectrometry analysis. Liquid chromatography was performed by using a reverse-phase C18 column (Zorbax C18, Agilent, 5 μm, 150x0.5 mm diameter column) and a Cogent diamond hydride column (MicroSolv, 4 μm, 150x2.1 mm diameter column) with a flow rate of 20 μl/min and 150 μl/min, respectively. For global profiling, samples were analyzed by using electrospray ionization time-of-flight mass spectrometry (Agilent 6538 TOF) with water/acetonitrile as mobile-phases A/B, each containing 0.1% formic acid. C18 sample analyses used either gradient 1 or gradient 2. Gradient 1 consisted of the following linear changes in mobile-phase B composition with time: 0 min: 10% B, 5 min: 10% B, 10 min: 40% B, 65 min: 98% B, 70 min: 98% B, 75 min: 10% B. Gradient 2 consisted of the following linear changes in mobile-phase B composition with time: 0 min: 10% B, 5 min: 10% B, 10 min: 40% B, 50 min: 98% B, 65 min: 98% B, 66 min: 10% B. Diamond hydride sample analyses started at 90% mobile-phase B with a 20-min linear gradient to 98% mobile-phase A. Sample groups were randomized from run to run, each separated by wash runs, to reduce error due to instrument variability and possible carryover. The samples were analyzed in both positive- and negative-ion mode.

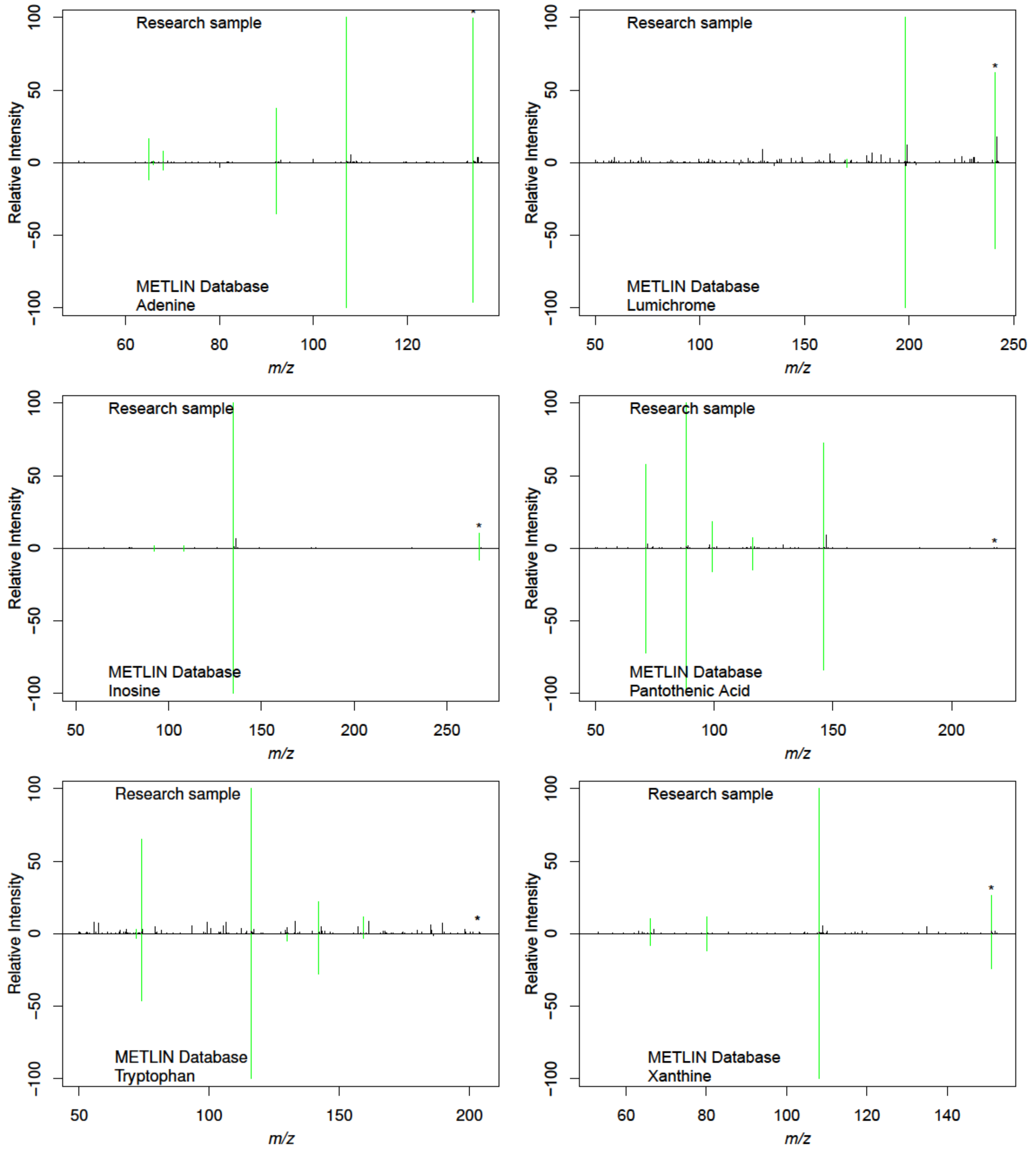
Representative examples of metabolites identified by using the automated workflow



Supplementary Figure 1: Representative examples of metabolites (human serum, positive ionization mode) identified by using the automated workflow for compounds with different physiochemical properties.

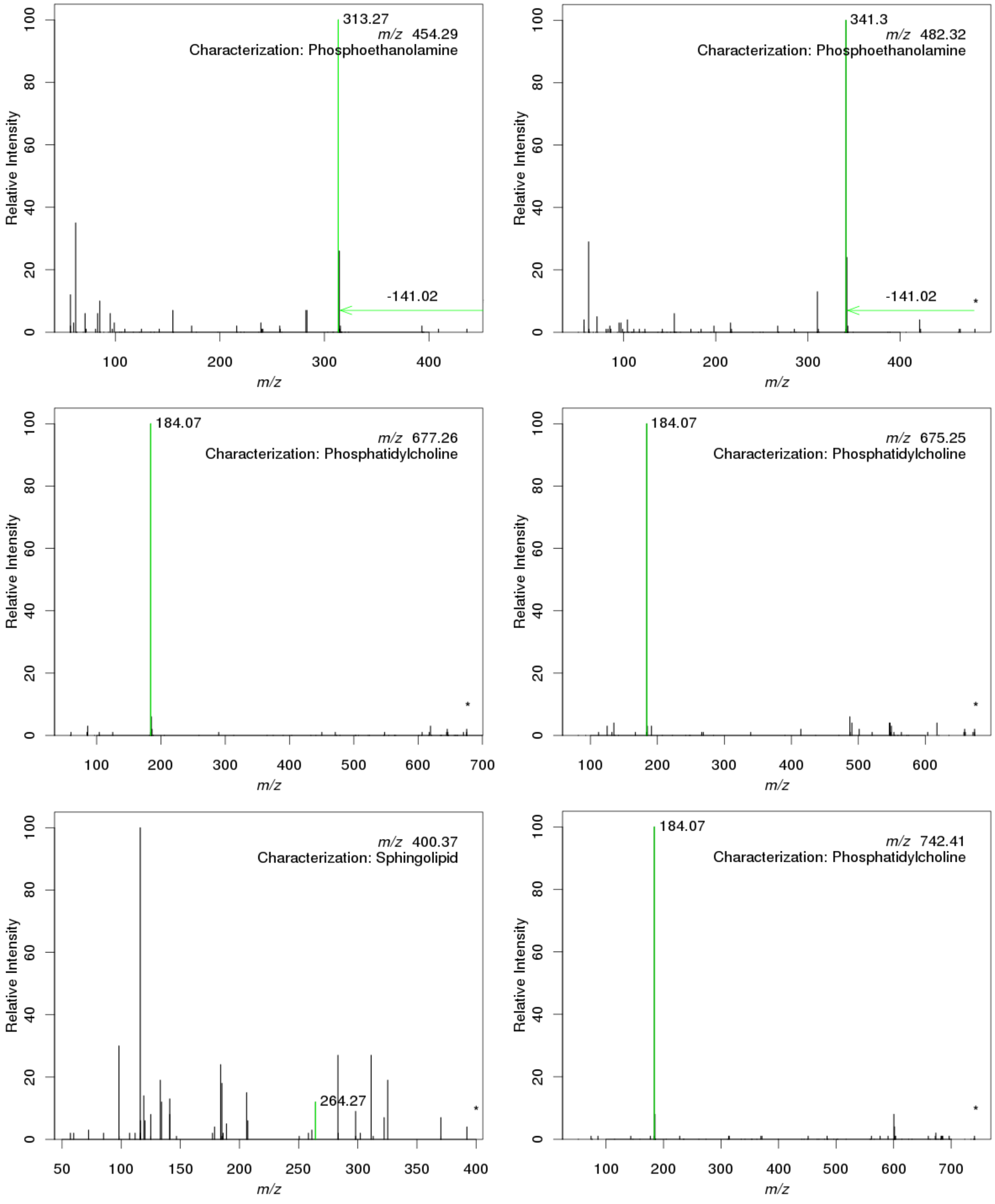


Supplementary Figure 2: Representative examples of metabolites (human serum, negative ionization mode) identified by using the automated workflow for compounds with different physiochemical properties.

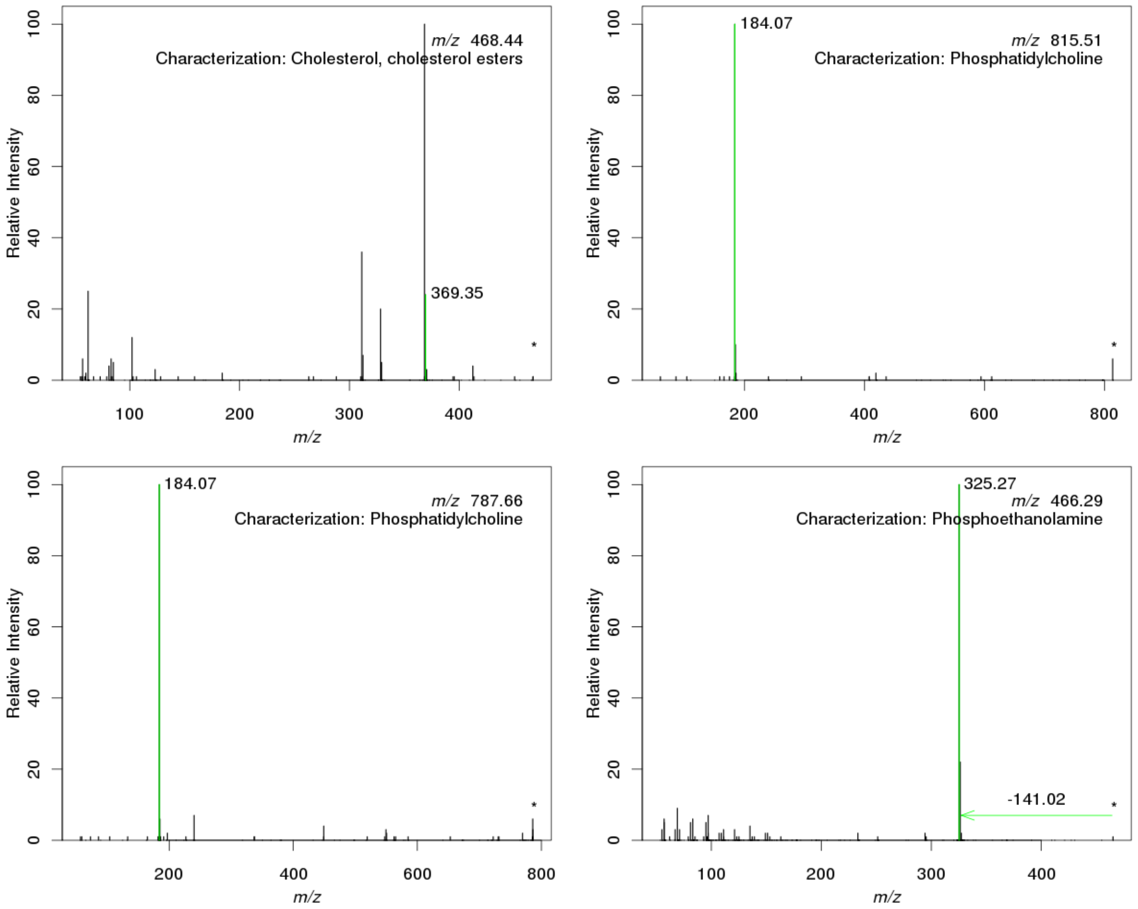


Supplementary Figure 3: Representative examples of metabolites (E.coli, negative ionization mode) identified by using the automated workflow for compounds with different physiochemical properties.

Representative examples for automated structural characterization by using the automated workflow



Supplementary Figure 4: Representative examples for automated structural characterization by using the automated workflow, shown for human serum in positive ionization mode.

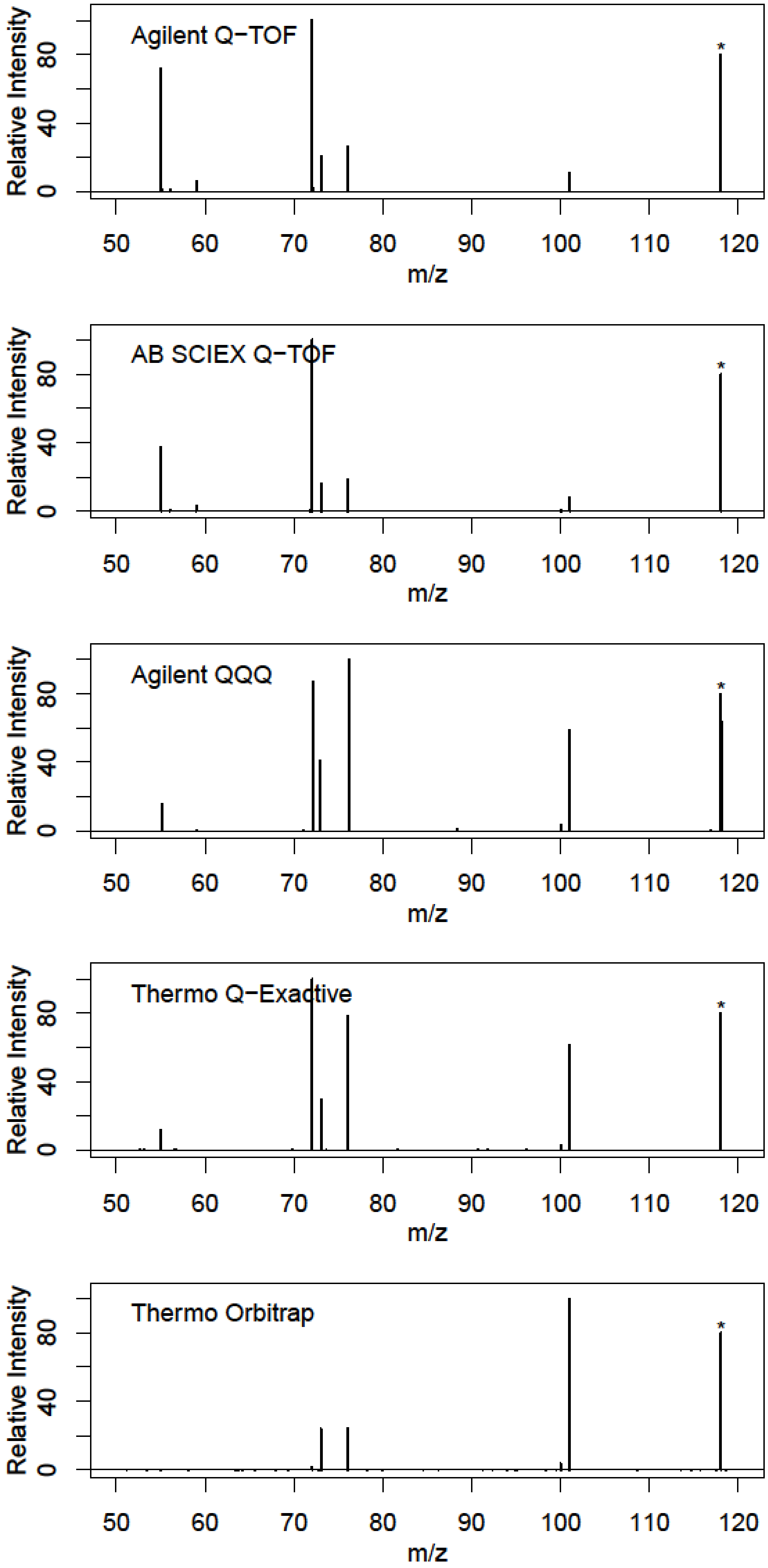


Supplementary Figure 5: Representative examples for automated structural characterization by using the automated workflow, shown for E.coli in positive ionization mode.

Comparison of METLIN Q-TOF data to data from other instruments

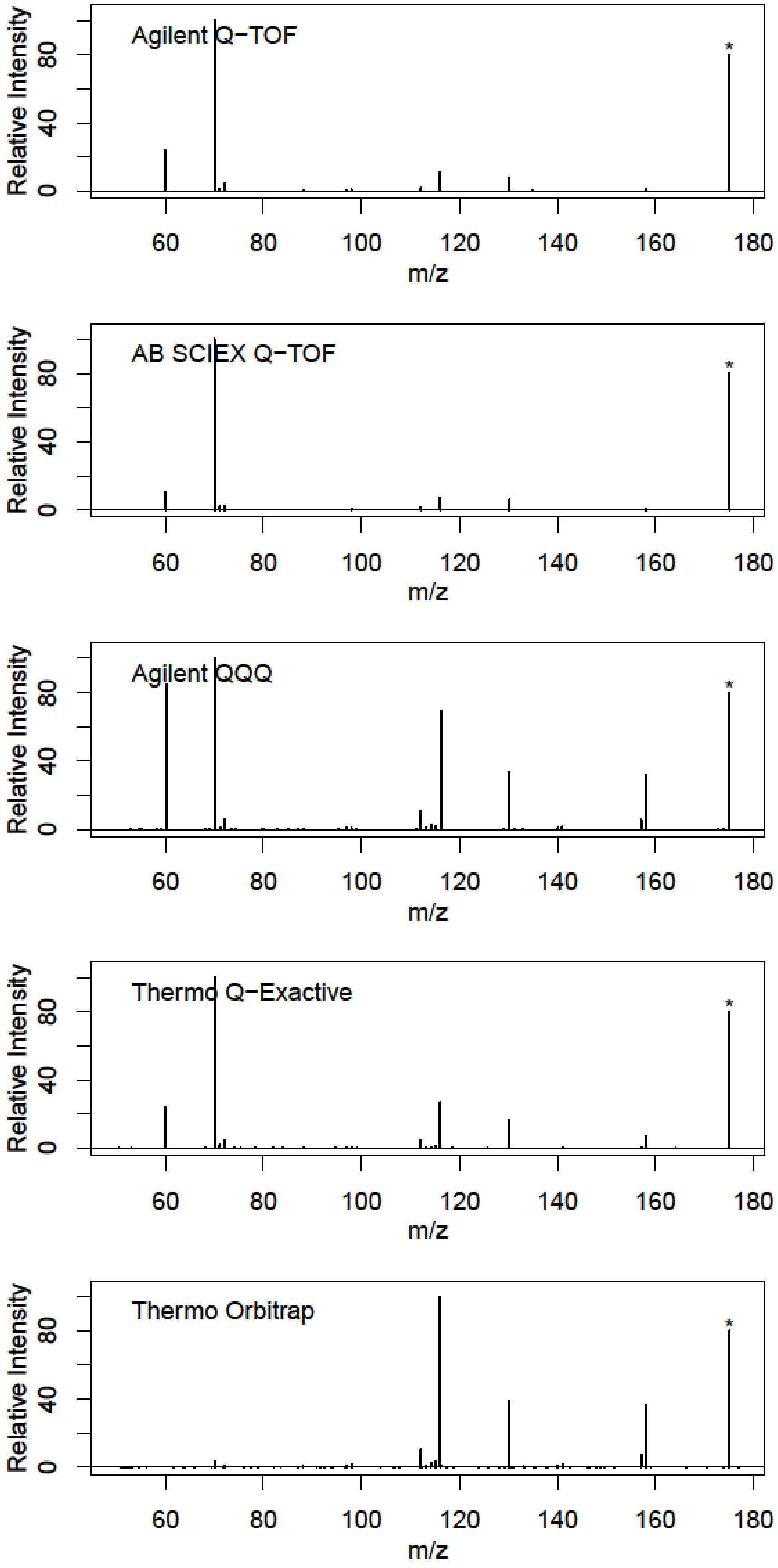
Compound name	Ionization mode	Agilent QTOF	AB SCIEX QTOF	Agilent QQQ	Thermo Orbitrap	Thermo Q-Exactive
Guanidineacetic acid	+	1	1	1	1	1
Arginine	+	1	1	1	1	1
Glutamate	+	1	1	3	1	1
Phenylalanine	+	1	1	1	1	1
Tryptophan	+	1	1	2	1	no signal
Melatonin	+	1	1	1	1	1
Homoserine	+	1	1	2	3	3
Caffeine	+	1	1	1	1	1
Nicotinuric acid	+	1	1	1	1	1
3-Methyl-L-histidine	+	1	1	1	1	1
5'-CMP	+	1	1	1	1	1
Curcumin	+	1	1	1	1	1
2-Aminopyrazine	+	1	1	1	1	1
Sunitinib	+	1	1	2	1	1
Warfarin	+	1	1	no signal	3	2
Cysteine-S-sulfate	+	1	1	1	1	1
L-NAME	+	1	1	1	1	1
Harmine	+	1	1	1	1	1
Ubiquinone (Q2)	+	1	1	2	1	1
5'-CMP	−	1	1	1	1	1
Traumatic Acid	−	1	1	1	1	1
2-Ethyl-2-Hydroxybutyric acid	−	1	1	1	1	1
D-Ala-D-Ala	−	1	1	1	no signal	1
Warfarin	−	1	1	1	no signal	2
Cysteine-S-sulfate	−	1	1	1	no signal	1
3,4-Dihydroxyphenyl ethanol	−	1	1	2	no signal	1
Harmine	−	1	1	1	no signal	1

Supplementary Table 1: Results of matching data from other instruments to the METLIN Q-TOF MS/MS database. Displayed are the results of the METLIN spectrum search[3] using mass spectra for 23 compounds that were acquired on five different instrument platforms: Agilent 6510 Q-TOF, AB SCIEX TripleTOF 5600, Agilent 6460 Triple Quad, Thermo Scientific LTQ Orbitrap Velos, and Thermo Scientific Q−Exactive mass spectrometers. The table displays the rank of the correct compound in the result list. Not counting the data from the Agilent 6510 Q-TOF (the instrument that was used to acquire all METLIN data and where the result is always the first hit) the correct compound was identified as the first hit for 90 out of the 101 spectra (89.1%) and structurally similar compounds were returned for 11 spectra (10.9%) . The individual spectra for each compound are shown in Supplemental Figures 6-32.



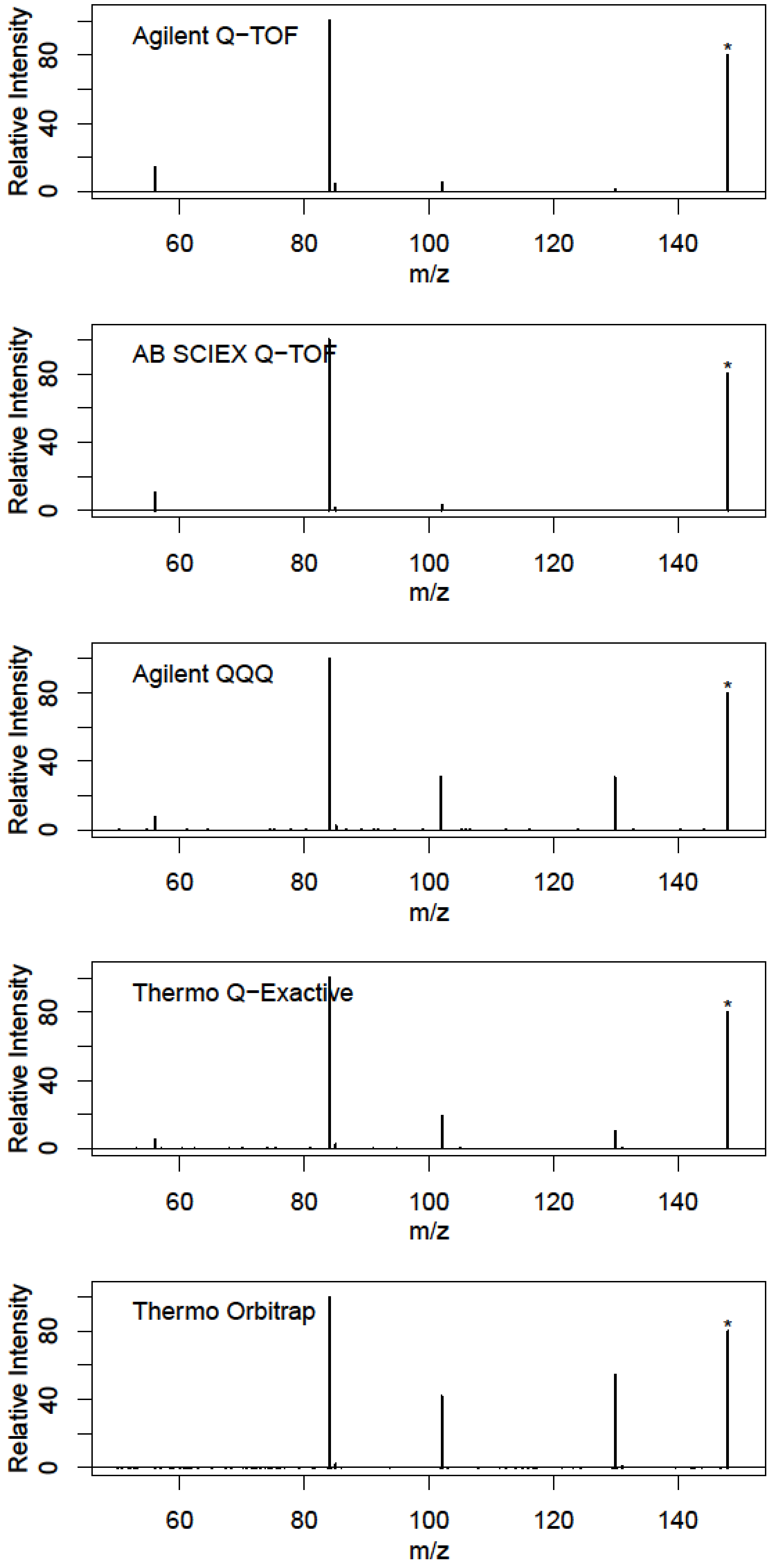
Supplementary Figure 6:

A comparison of MS^2 data for guanidineacetic acid stored in METLIN to MS^2 data for guanidineacetic acid generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in positive ionization mode.



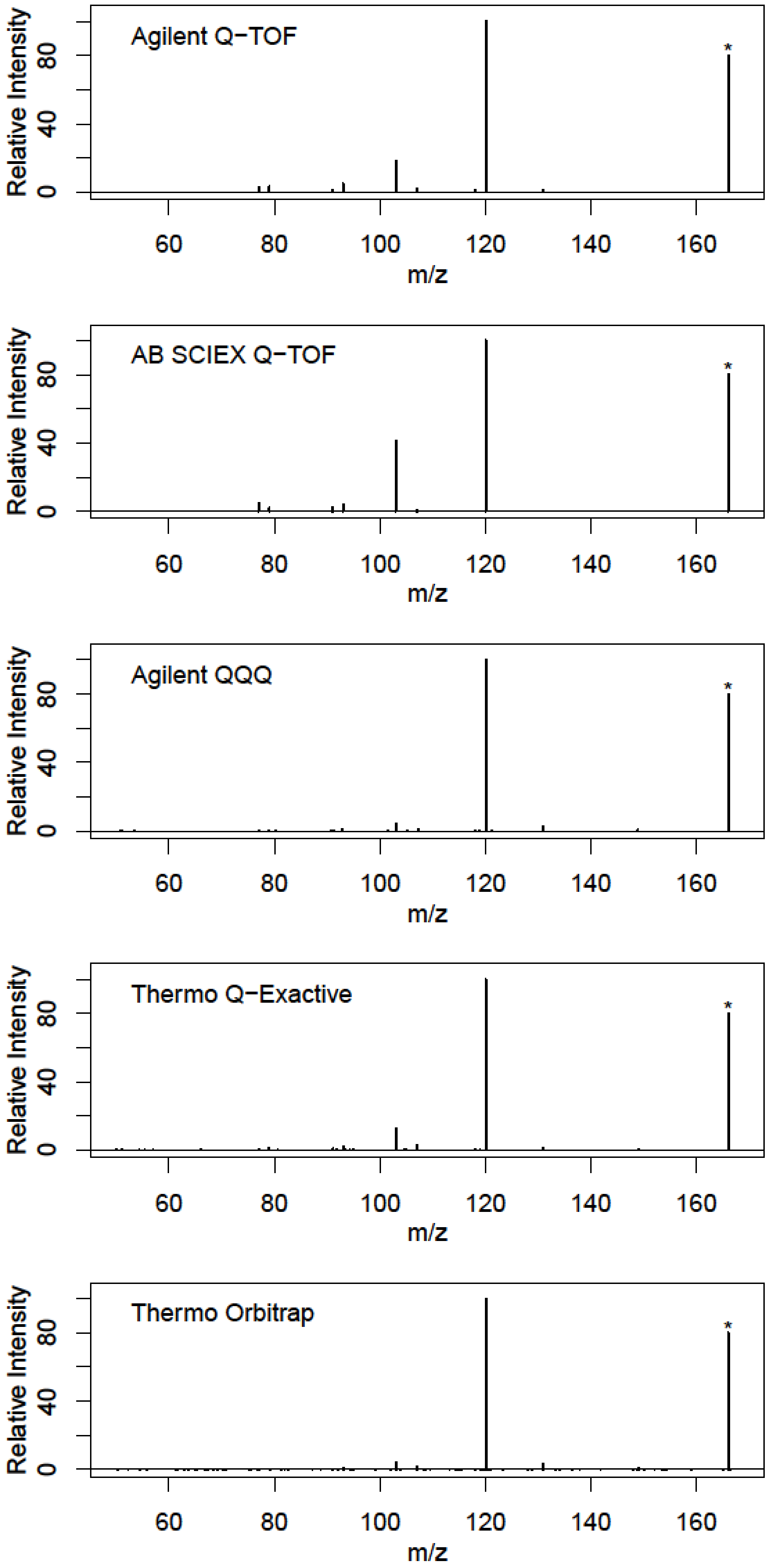
Supplementary Figure 7:

A comparison of MS^2 data for arginine stored in METLIN to MS^2 data for arginine generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in positive ionization mode.



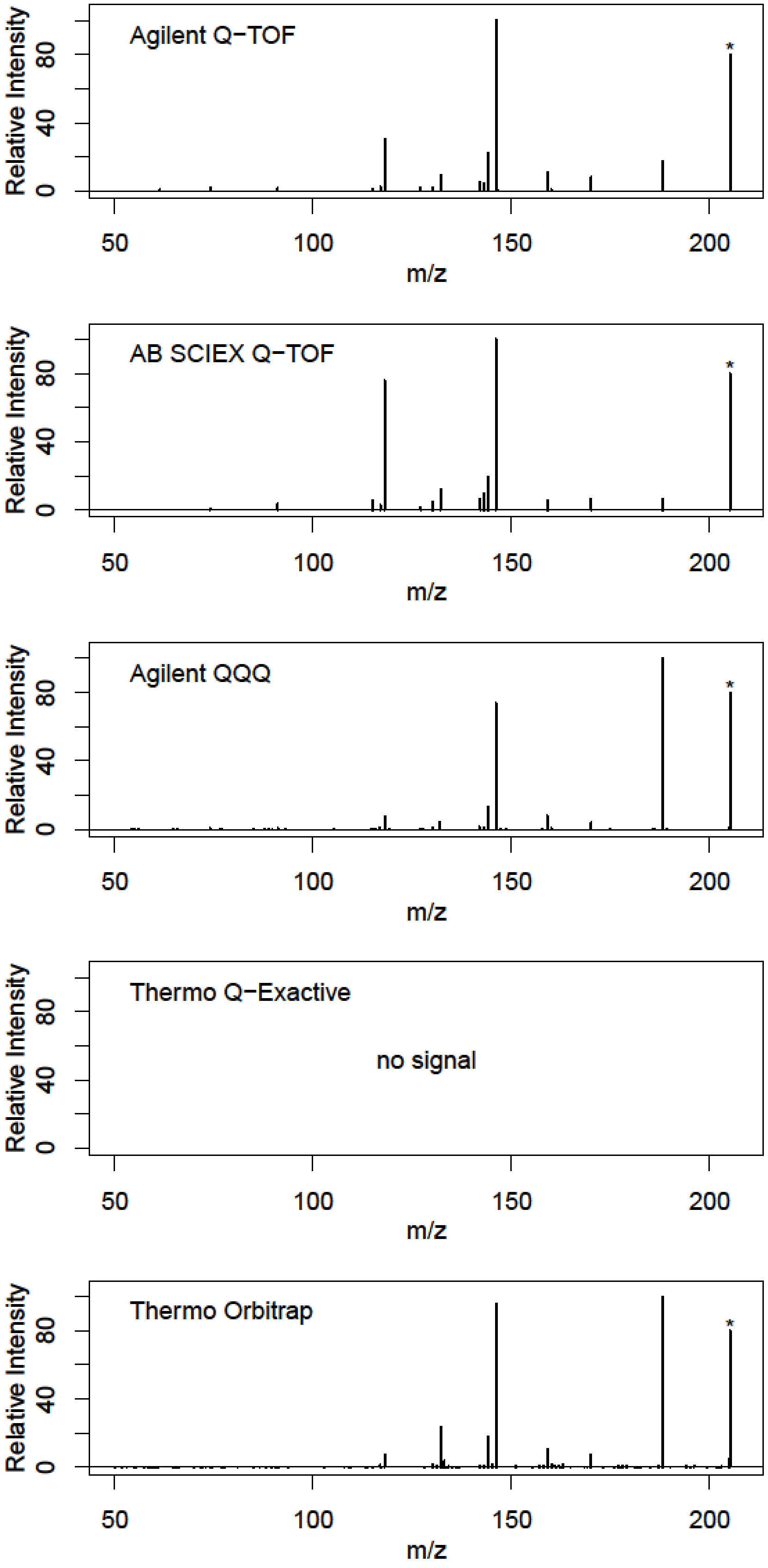
Supplementary Figure 8:

A comparison of MS^2 data for glutamate stored in METLIN to MS^2 data for glutamate generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in positive ionization mode.



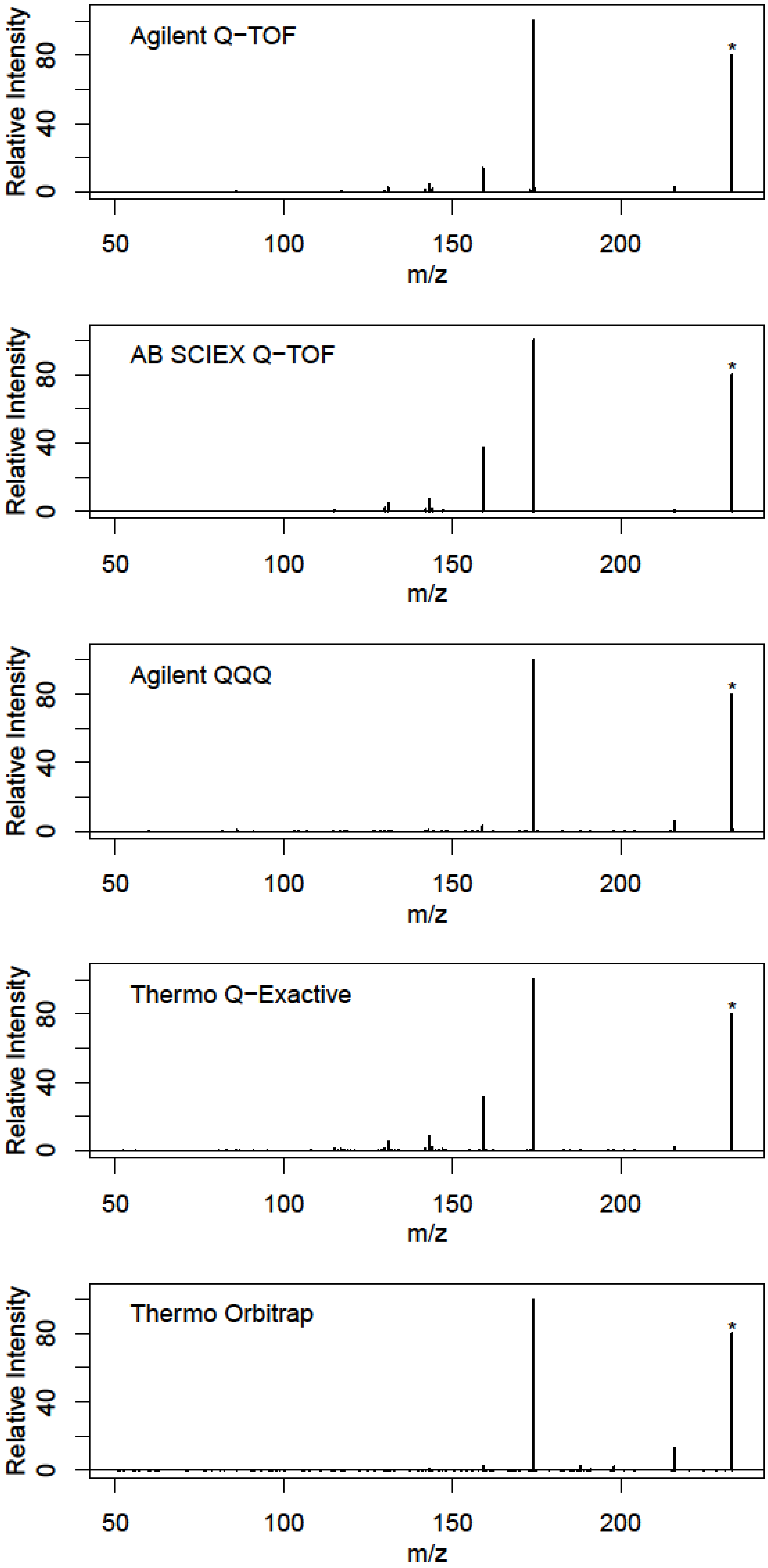
Supplementary Figure 9:

A comparison of MS^2 data for phenylalanine stored in METLIN to MS^2 data for phenylalanine generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in positive ionization mode.



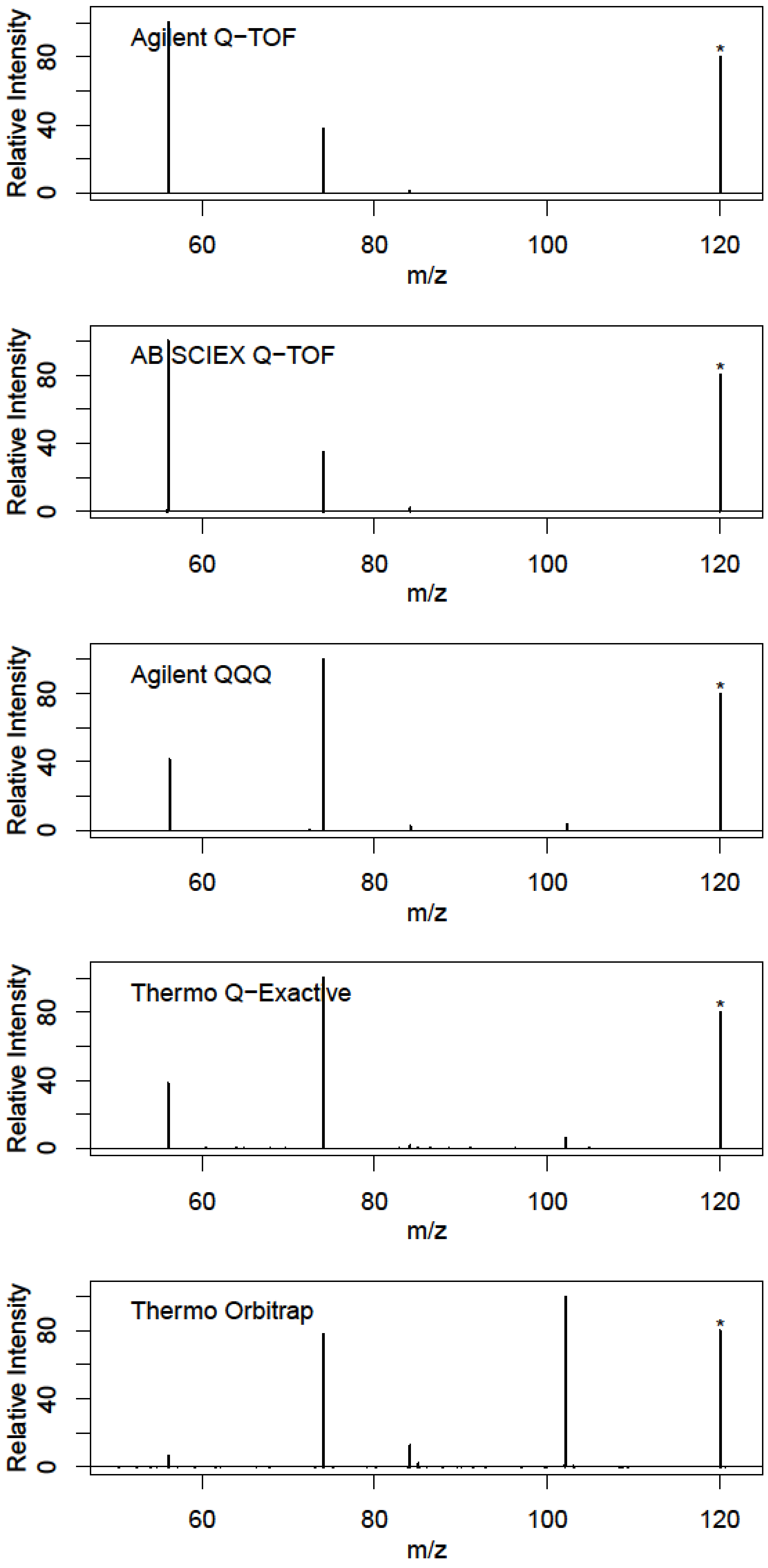
Supplementary Figure 10:

A comparison of MS^2 data for tryptophan stored in METLIN to MS^2 data for tryptophan generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in positive ionization mode.



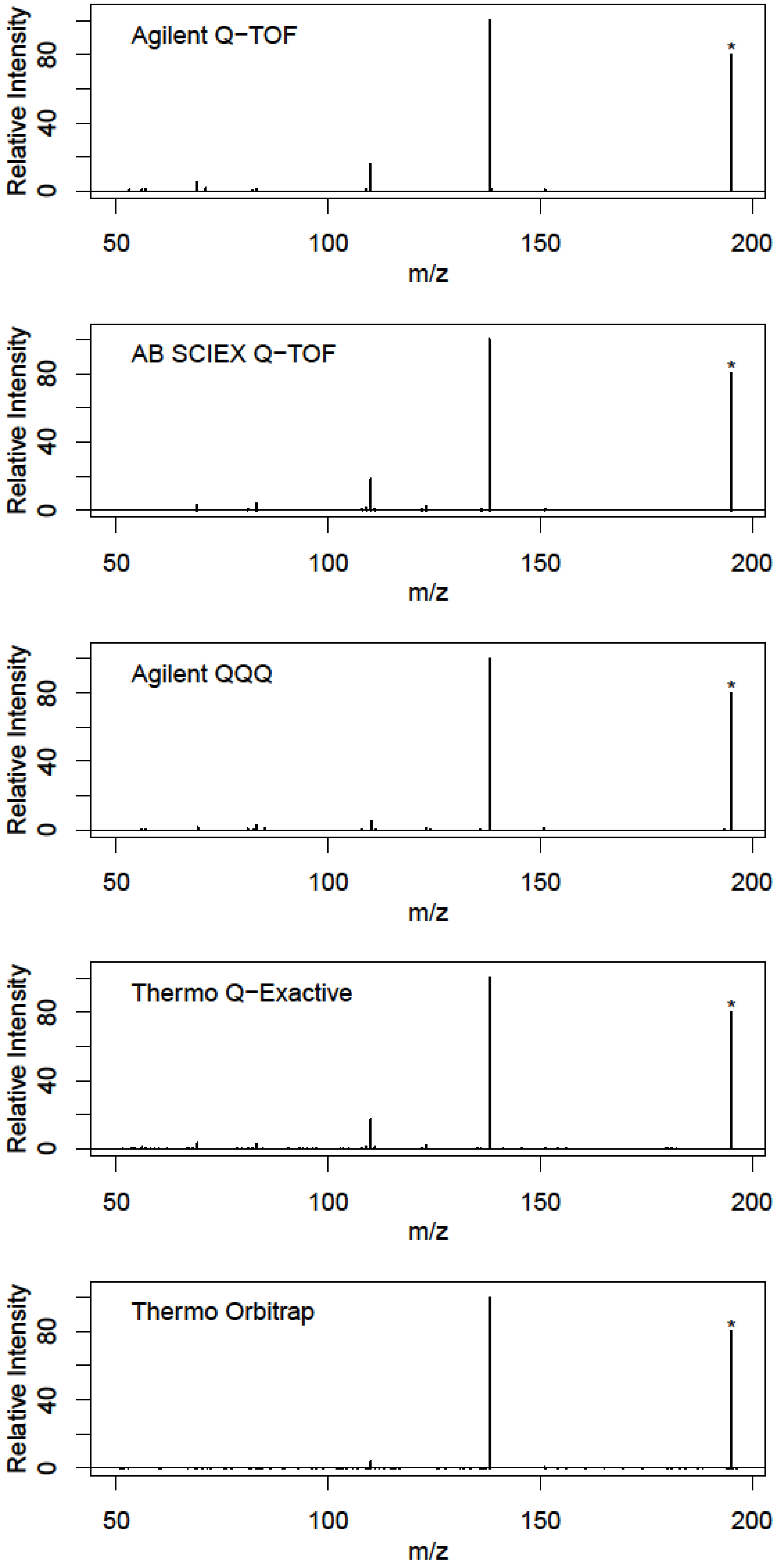
Supplementary Figure 11:

A comparison of MS^2 data for melatonin stored in METLIN to MS^2 data for melatonin generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in positive ionization mode.



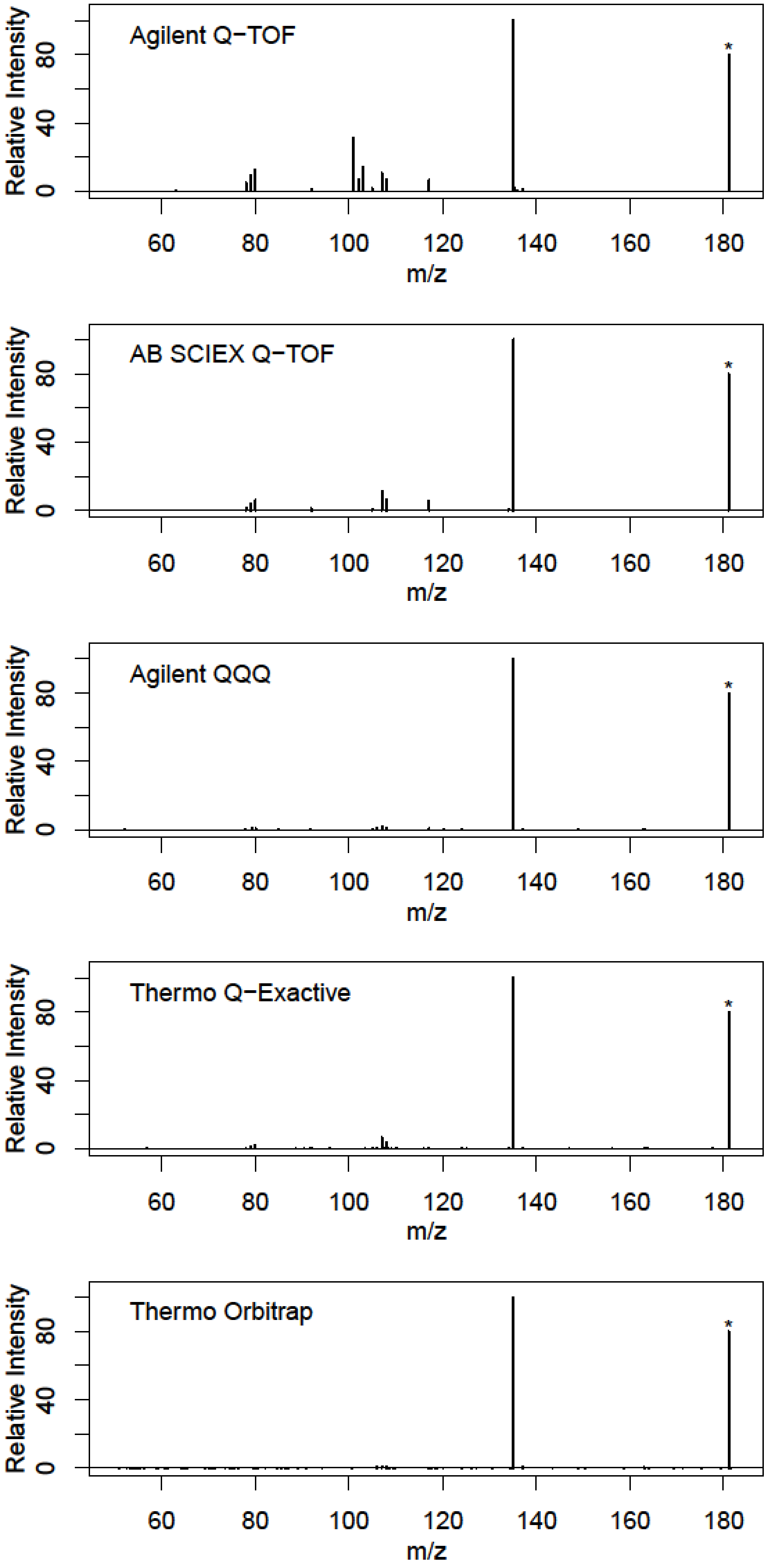
Supplementary Figure 12:

A comparison of MS^2 data for homoserine stored in METLIN to MS^2 data for homoserine generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in positive ionization mode.



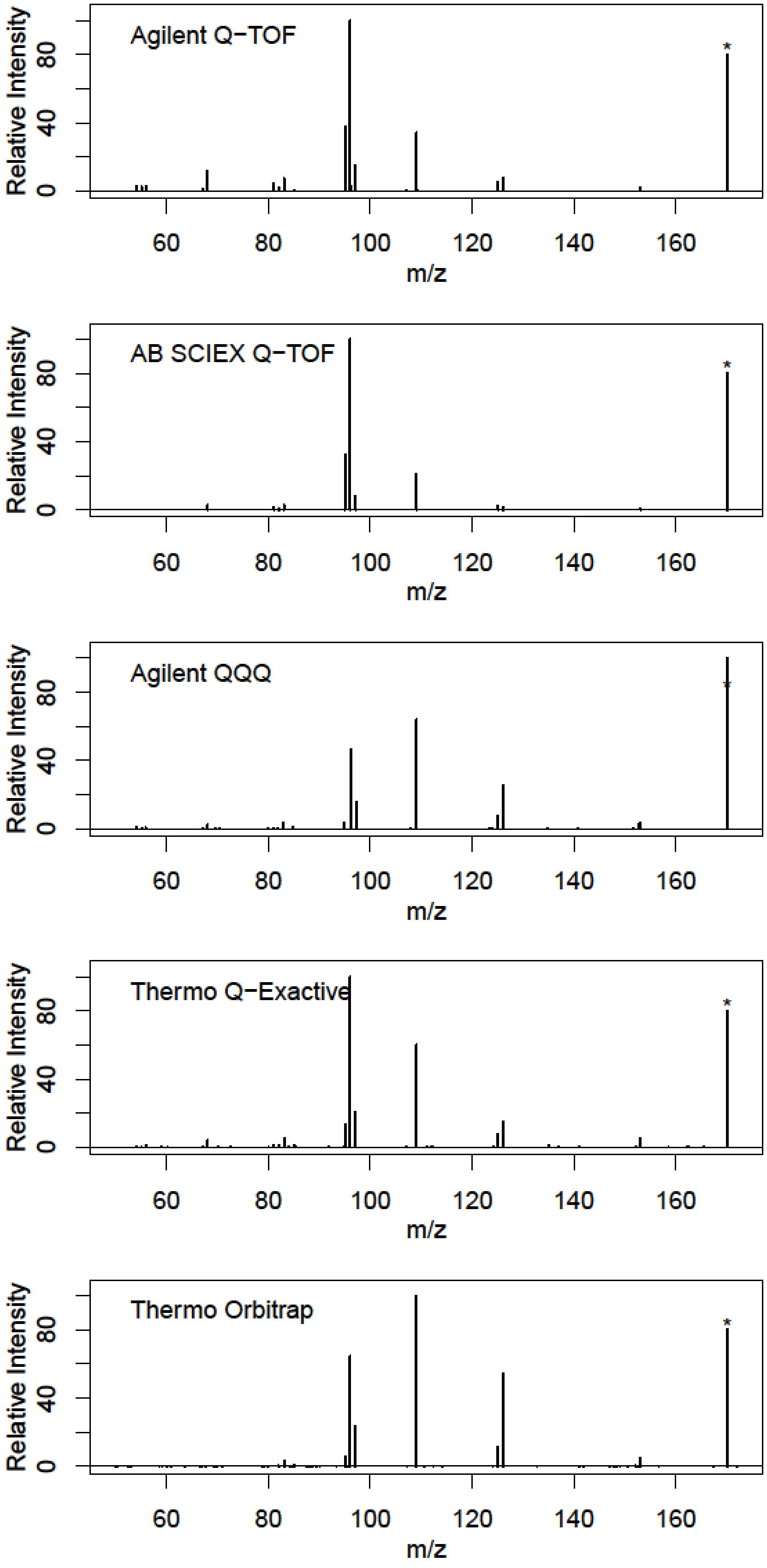
Supplementary Figure 13:

A comparison of MS^2 data for caffeine stored in METLIN to MS^2 data for caffeine generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in positive ionization mode.



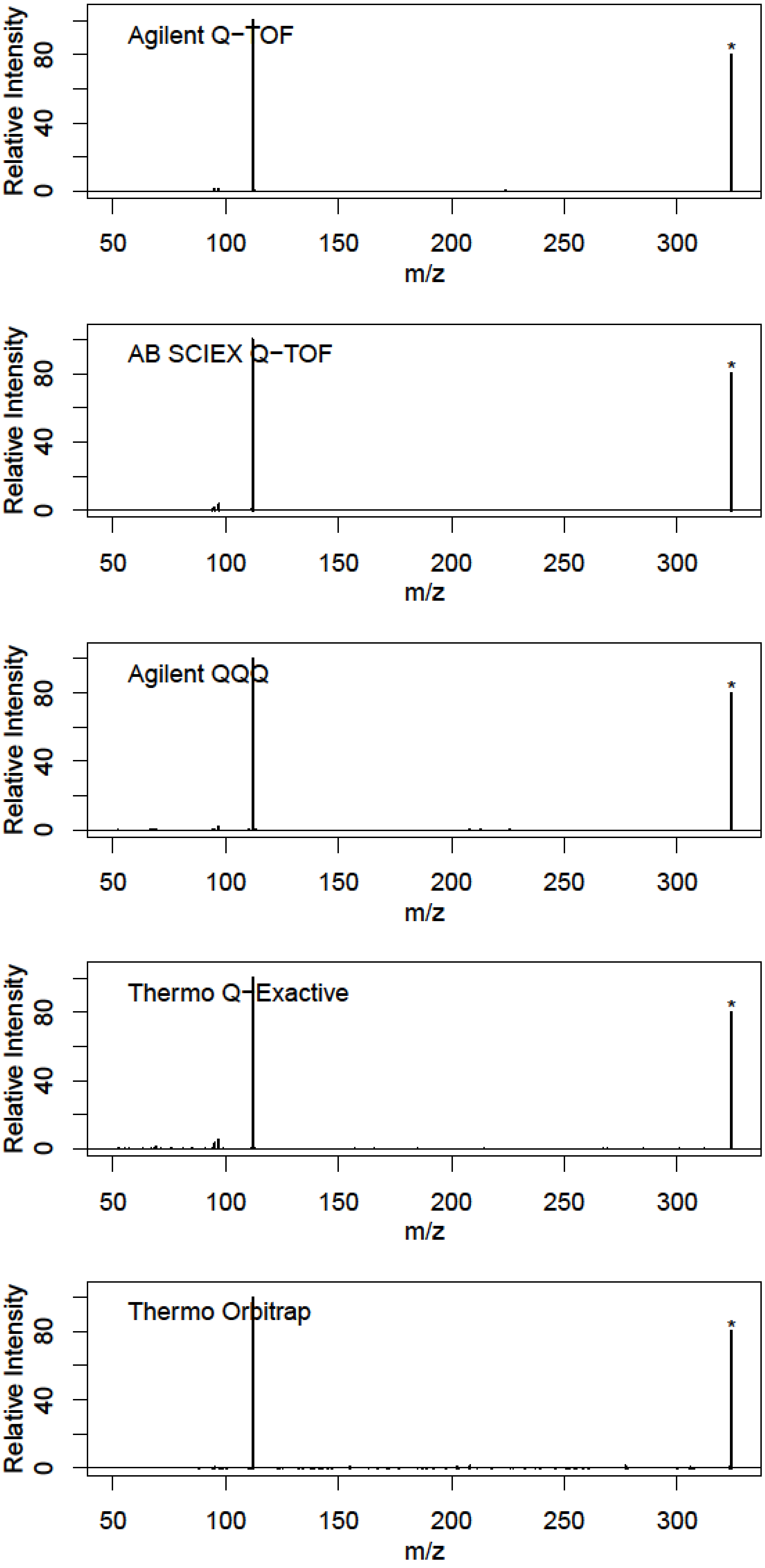
Supplementary Figure 14:

A comparison of MS^2 data for nicotinuric acid stored in METLIN to MS^2 data for nicotinuric acid generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in positive ionization mode.



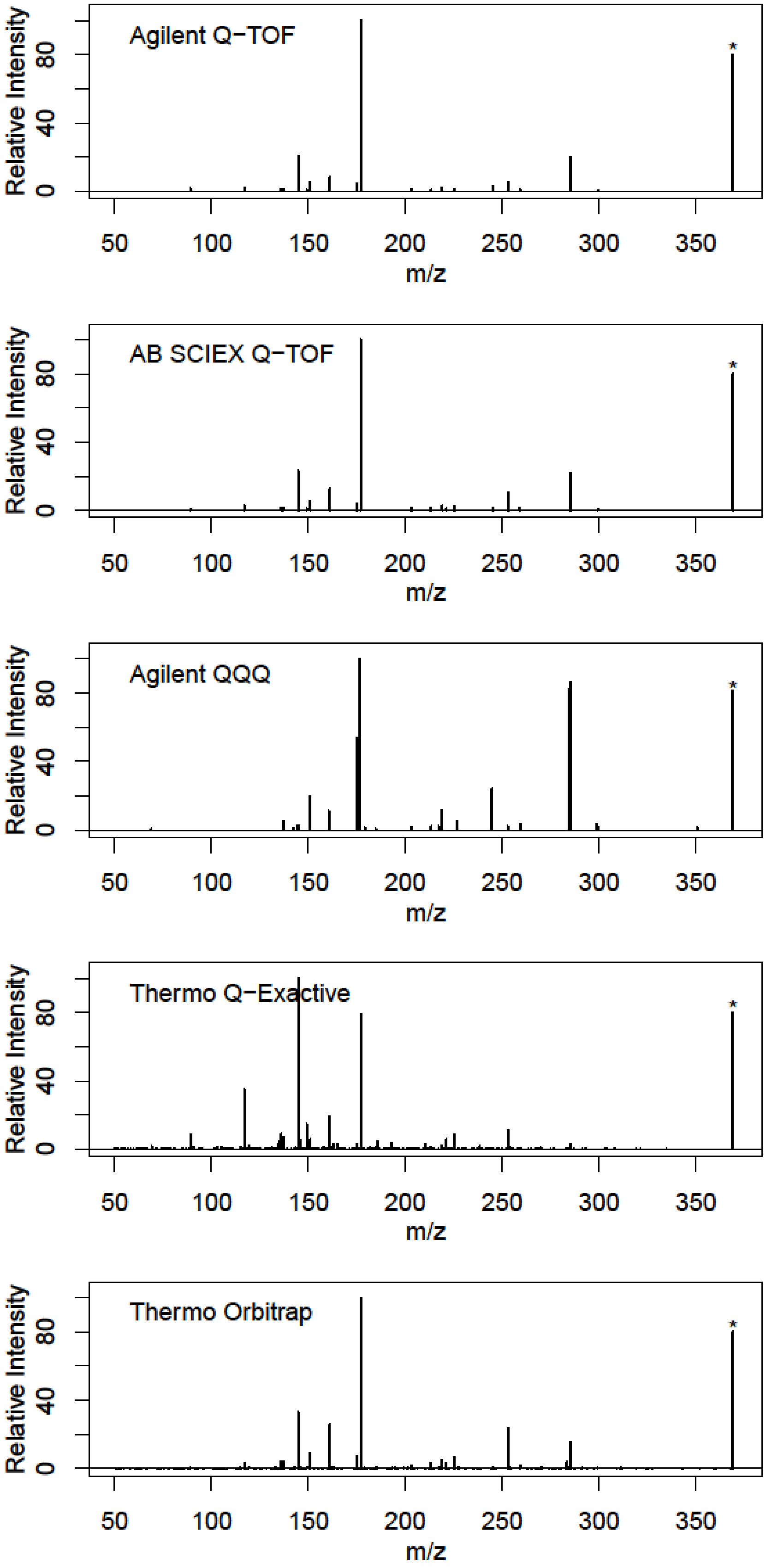
Supplementary Figure 15:

A comparison of MS^2 data for 3-methyl-L-histidine stored in METLIN to MS^2 data for 3-methyl-L-histidine generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in positive ionization mode.



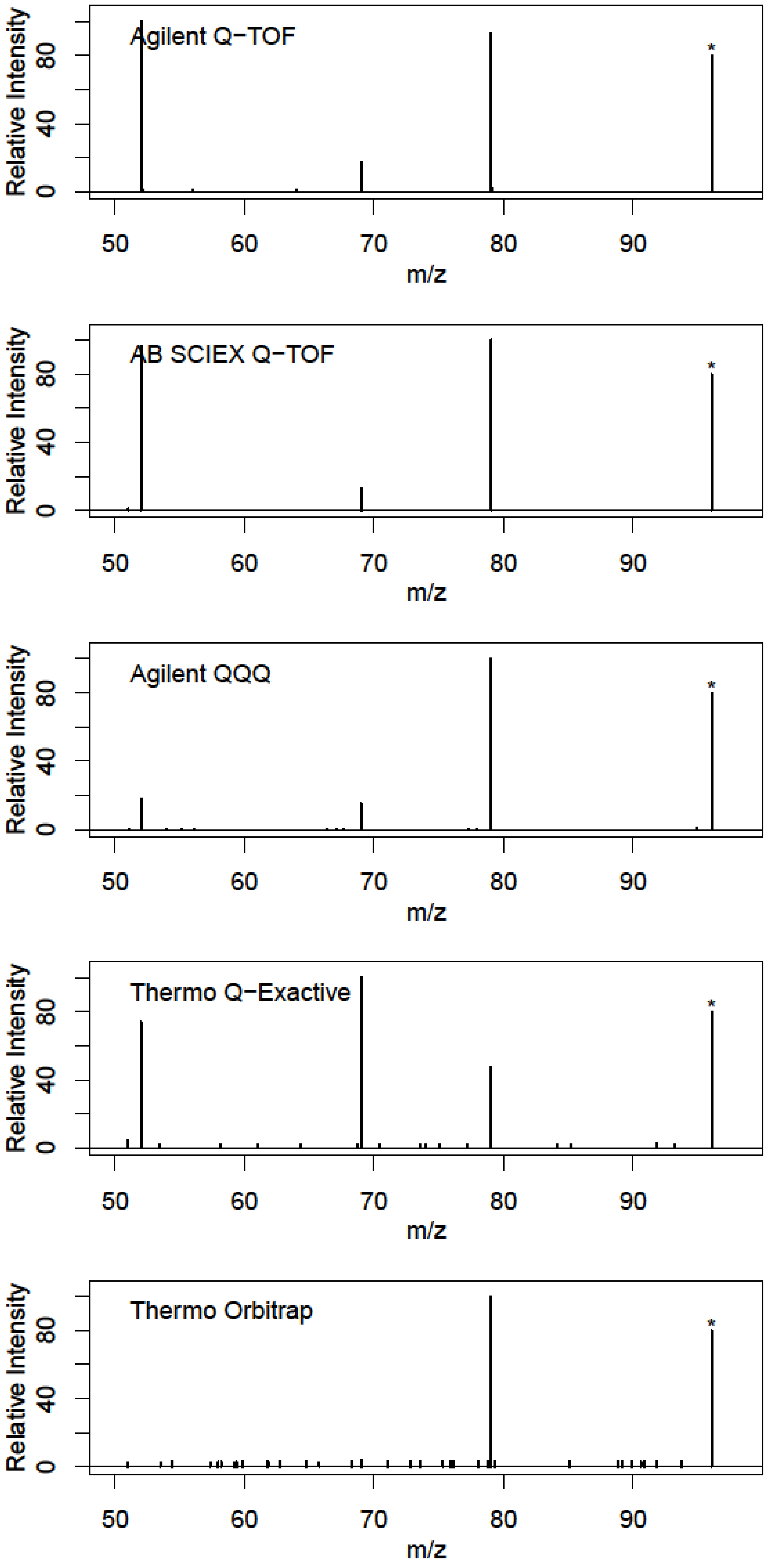
Supplementary Figure 16:

A comparison of MS^2 data for 5'-CMP stored in METLIN to MS^2 data for 5'-CMP generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in positive ionization mode.



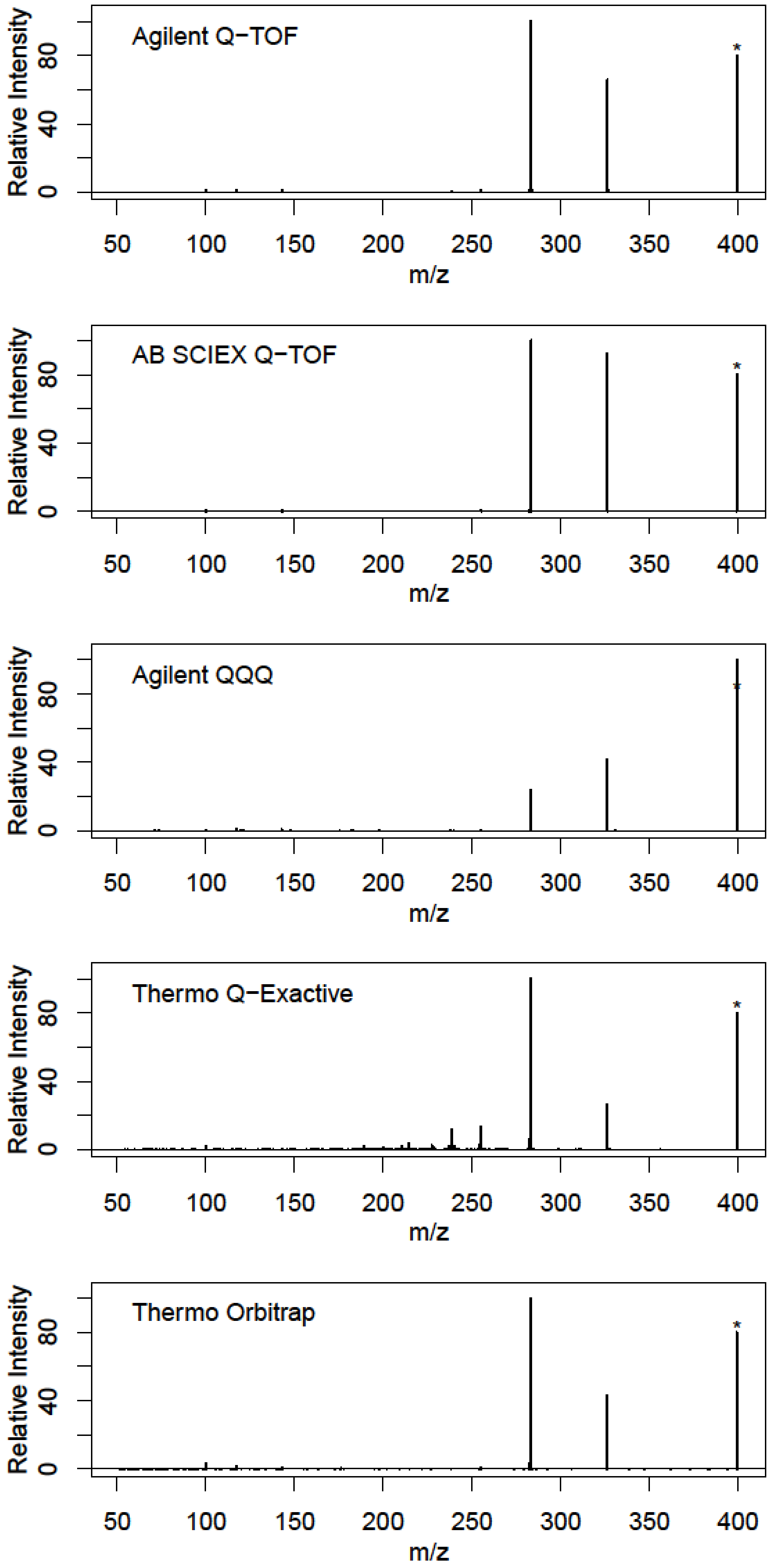
Supplementary Figure 17:

A comparison of MS^2 data for curcumin stored in METLIN to MS^2 data for curcumin generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in positive ionization mode.



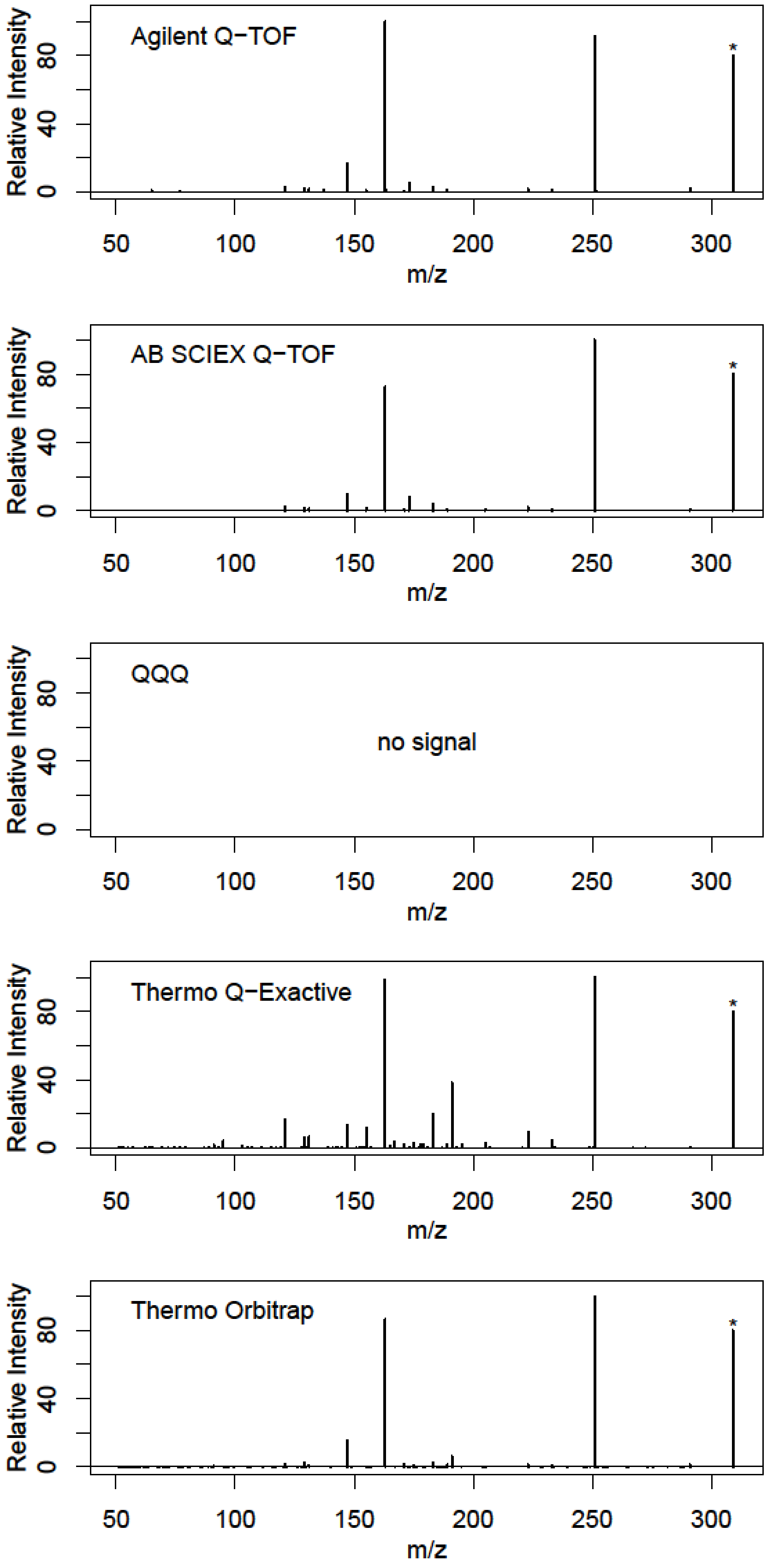
Supplementary Figure 18:

A comparison of MS^2 data for 2-aminopyrazine stored in METLIN to MS^2 data for 2-aminopyrazine generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in positive ionization mode.



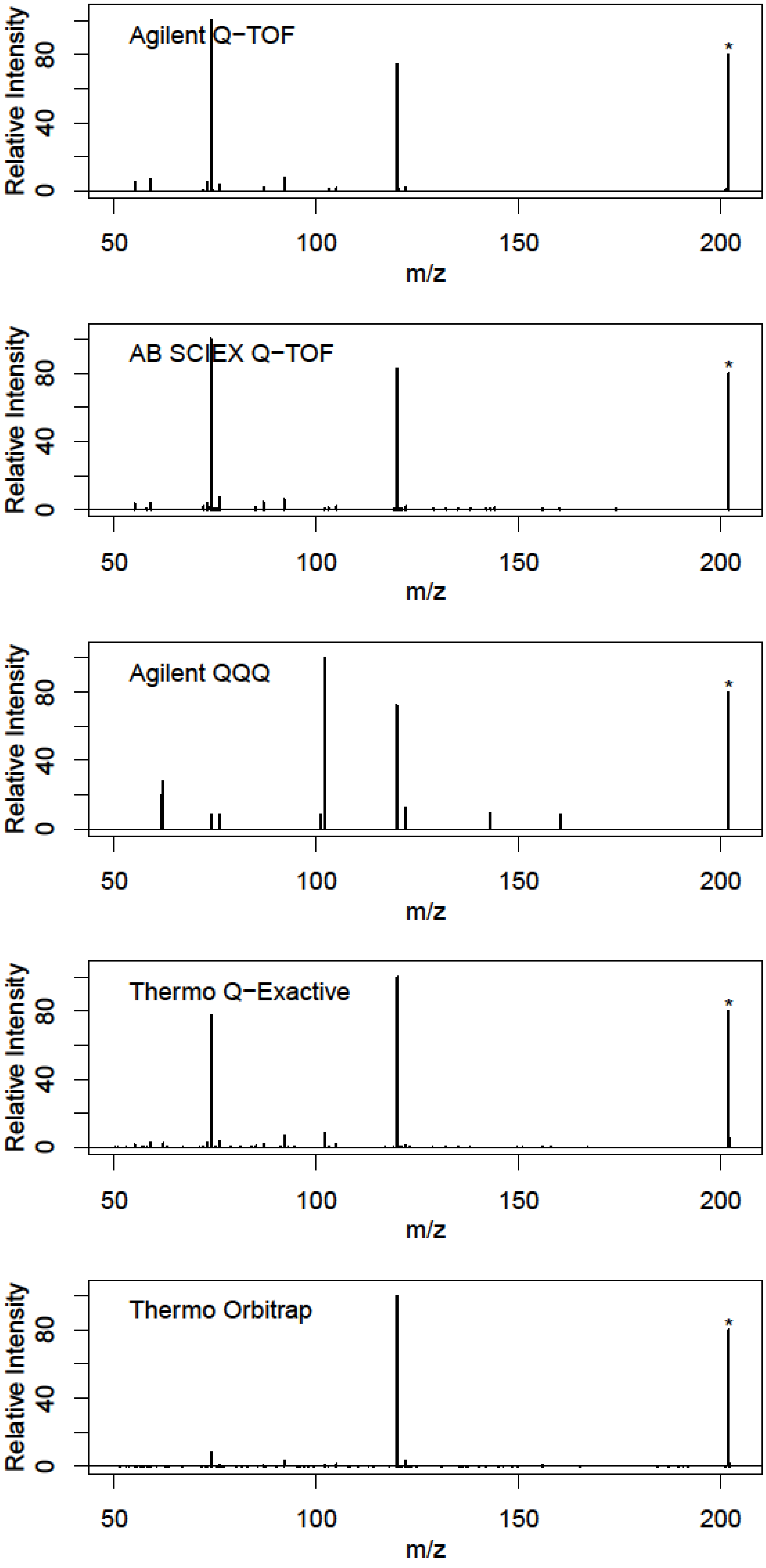
Supplementary Figure 19:

A comparison of MS^2 data for sunitinib stored in METLIN to MS^2 data for sunitinib generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in positive ionization mode.



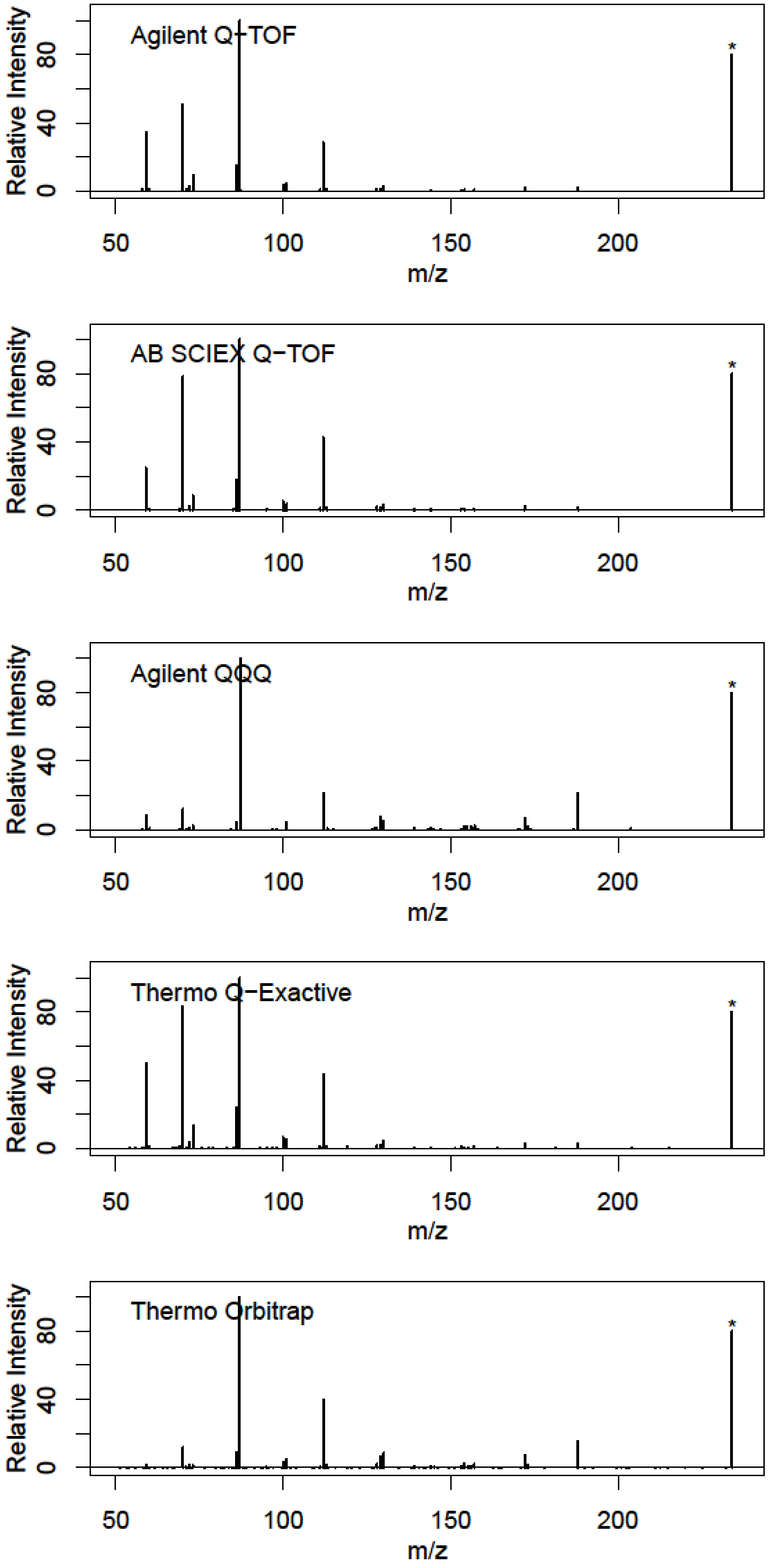
Supplementary Figure 20:

A comparison of MS^2 data for warfarin stored in METLIN to MS^2 data for warfarin generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in positive ionization mode.



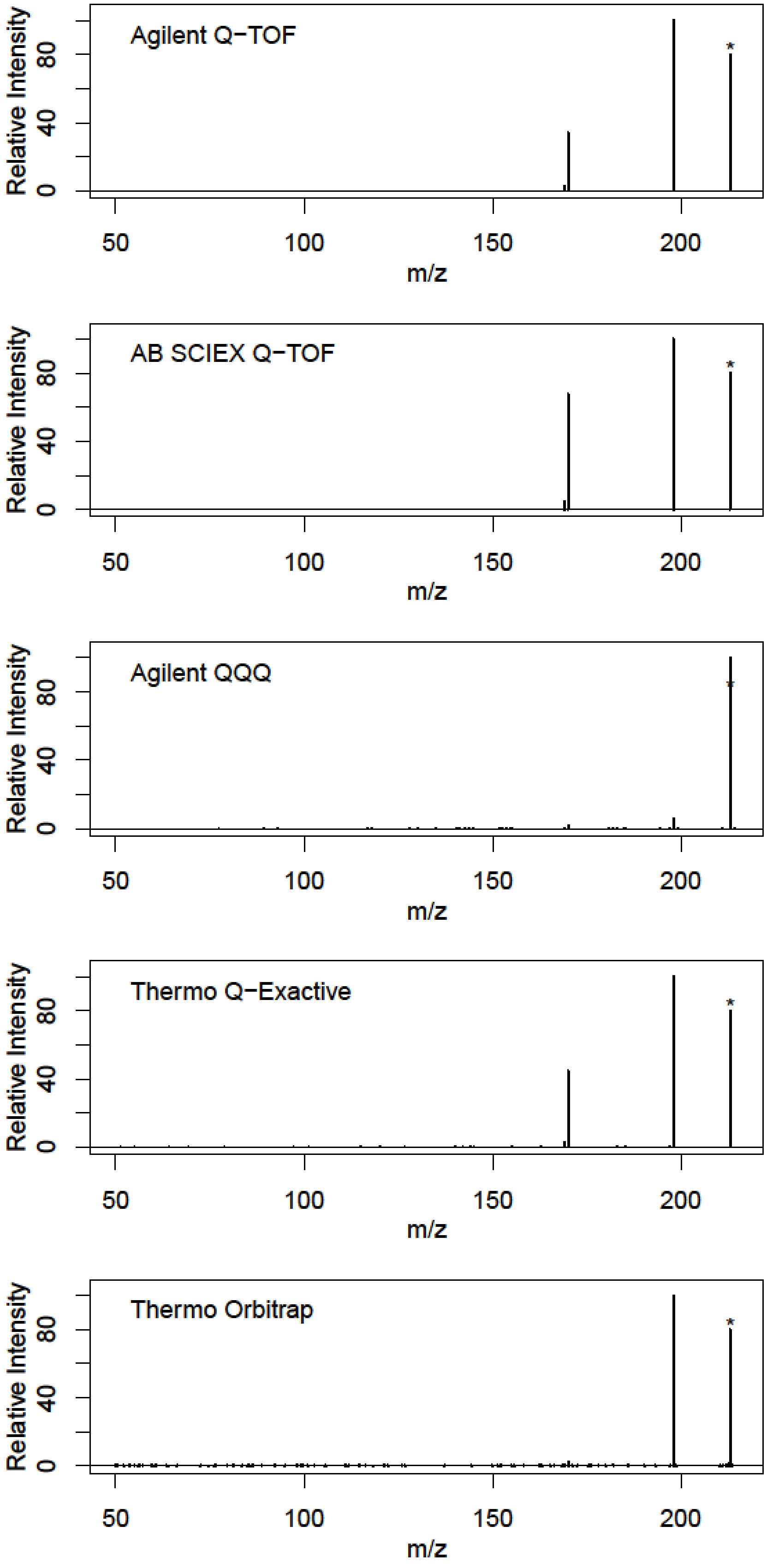
Supplementary Figure 21:

A comparison of MS^2 data for cysteine-S-sulfate stored in METLIN to MS^2 data for cysteine-S-sulfate generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in positive ionization mode.



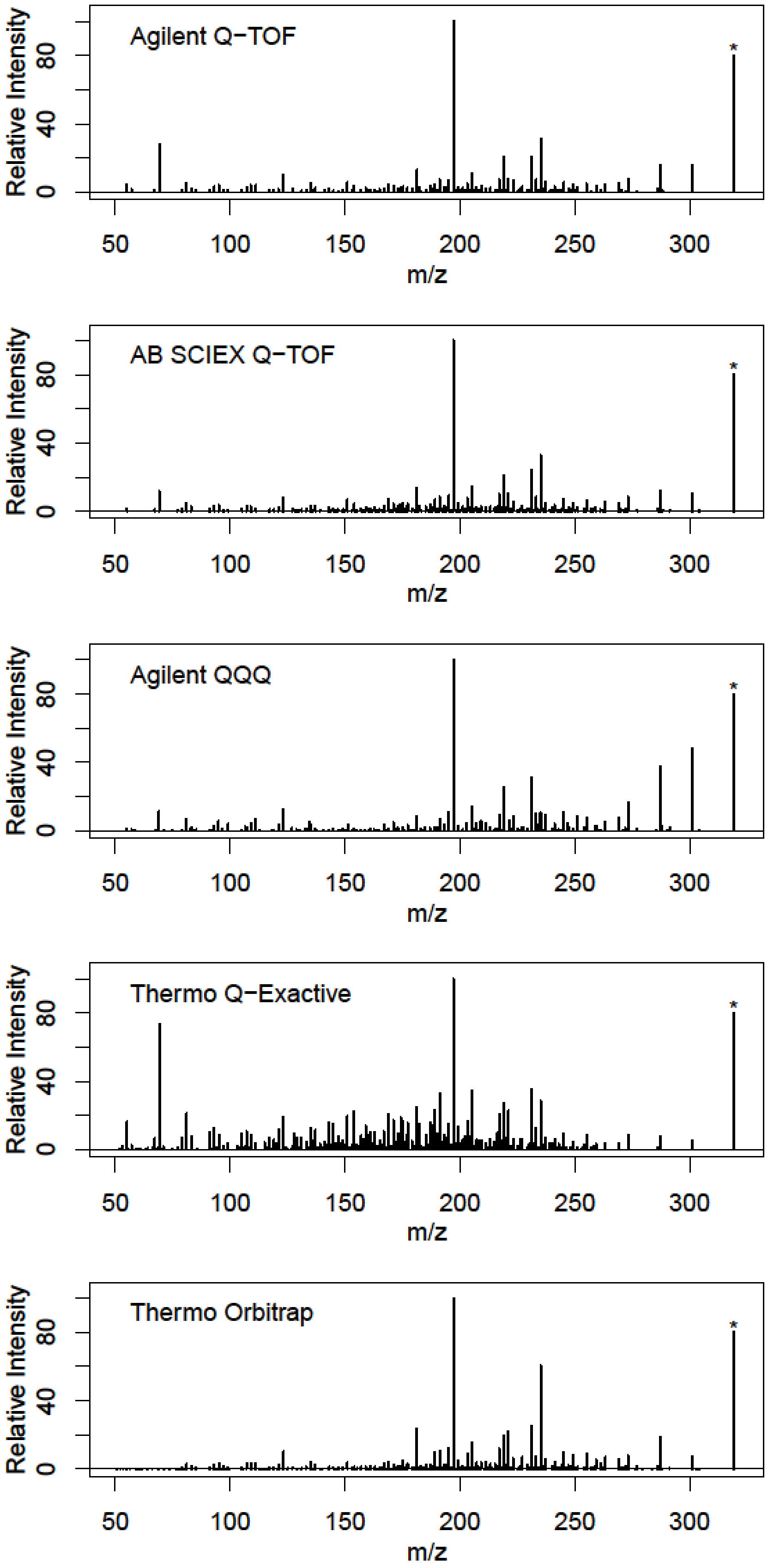
Supplementary Figure 22:

A comparison of MS^2 data for L-NAME stored in METLIN to MS^2 data for L-NAME generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in positive ionization mode.



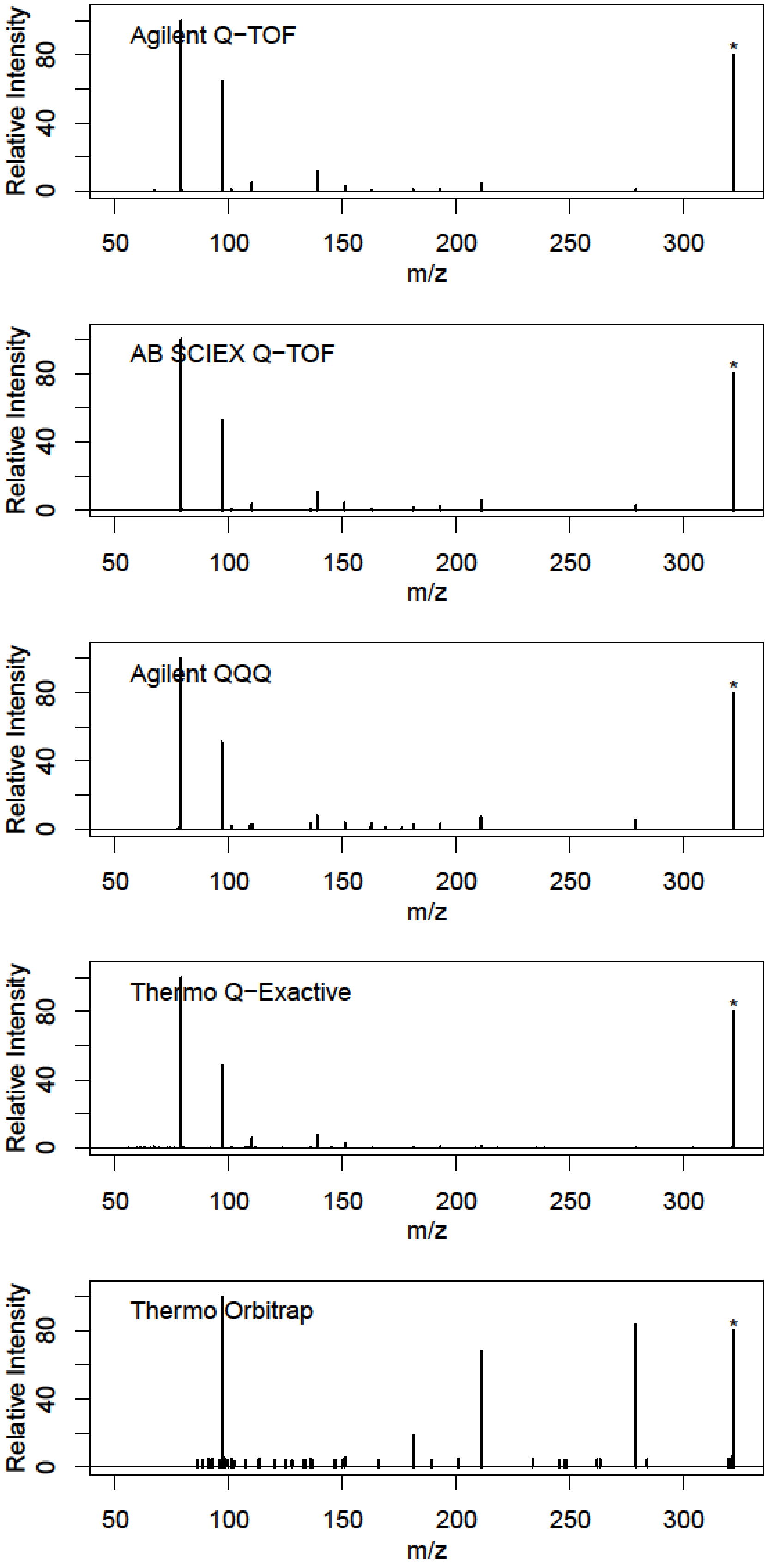
Supplementary Figure 23:

A comparison of MS^2 data for harmine stored in METLIN to MS^2 data for harmine generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in positive ionization mode.



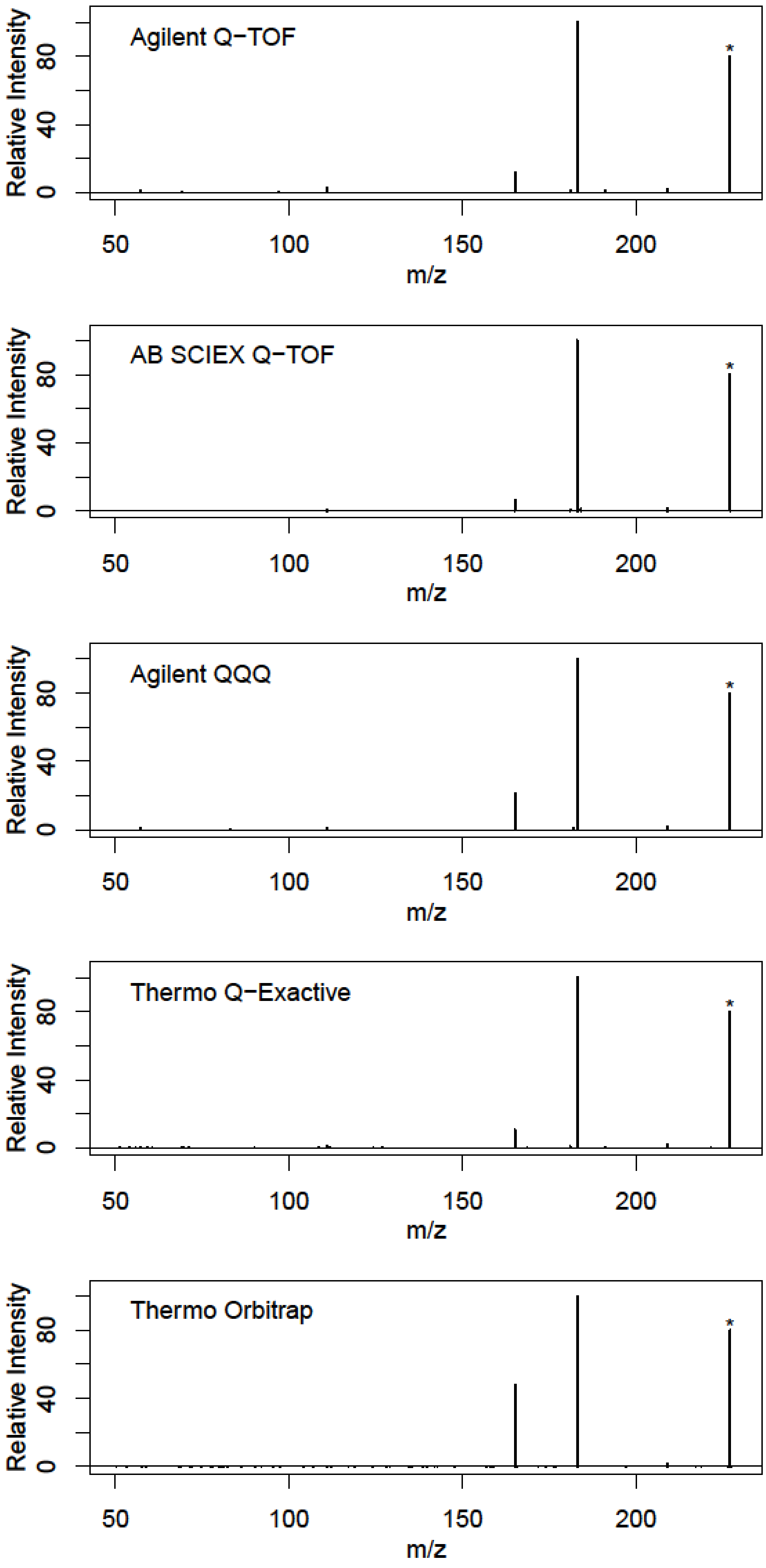
Supplementary Figure 24:

A comparison of MS^2 data for ubiquinone (Q2) stored in METLIN to MS^2 data for ubiquinone (Q2) generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in positive ionization mode.



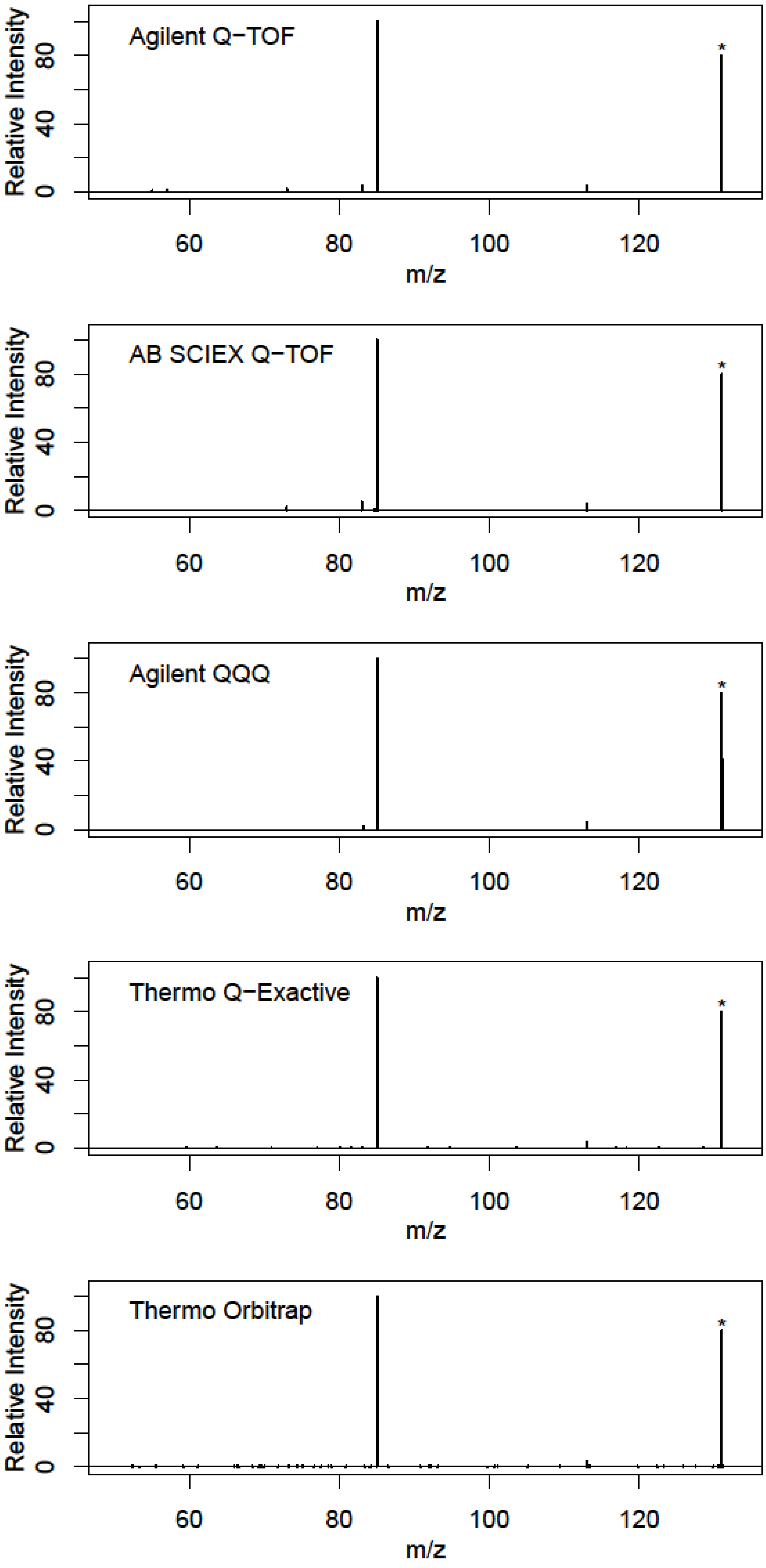
Supplementary Figure 25:

A comparison of MS^2 data for 5'-CMP stored in METLIN to MS^2 data for 5'-CMP generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in negative ionization mode.



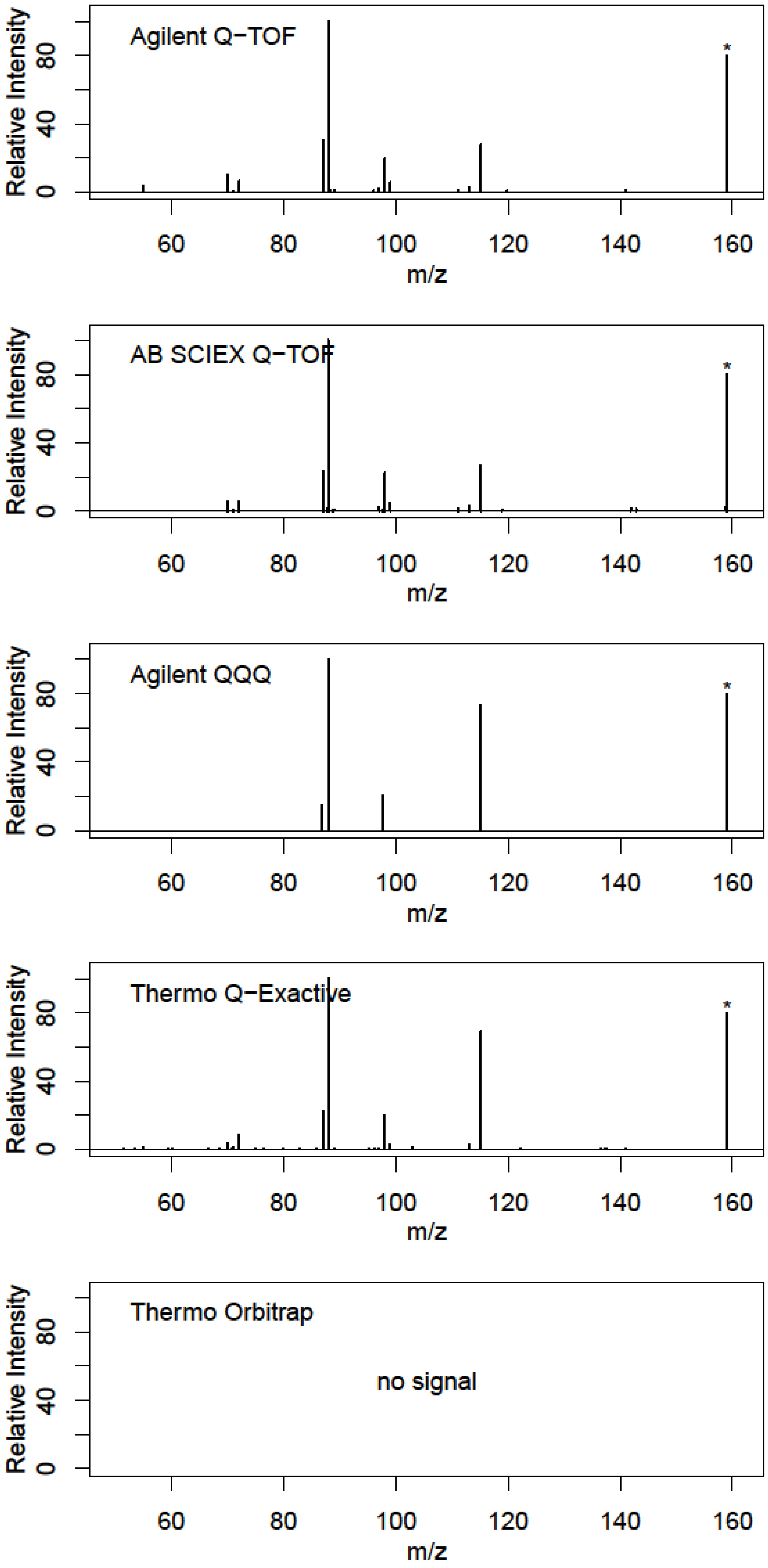
Supplementary Figure 26:

A comparison of MS^2 data for traumatic acid stored in METLIN to MS^2 data for traumatic acid generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in negative ionization mode.



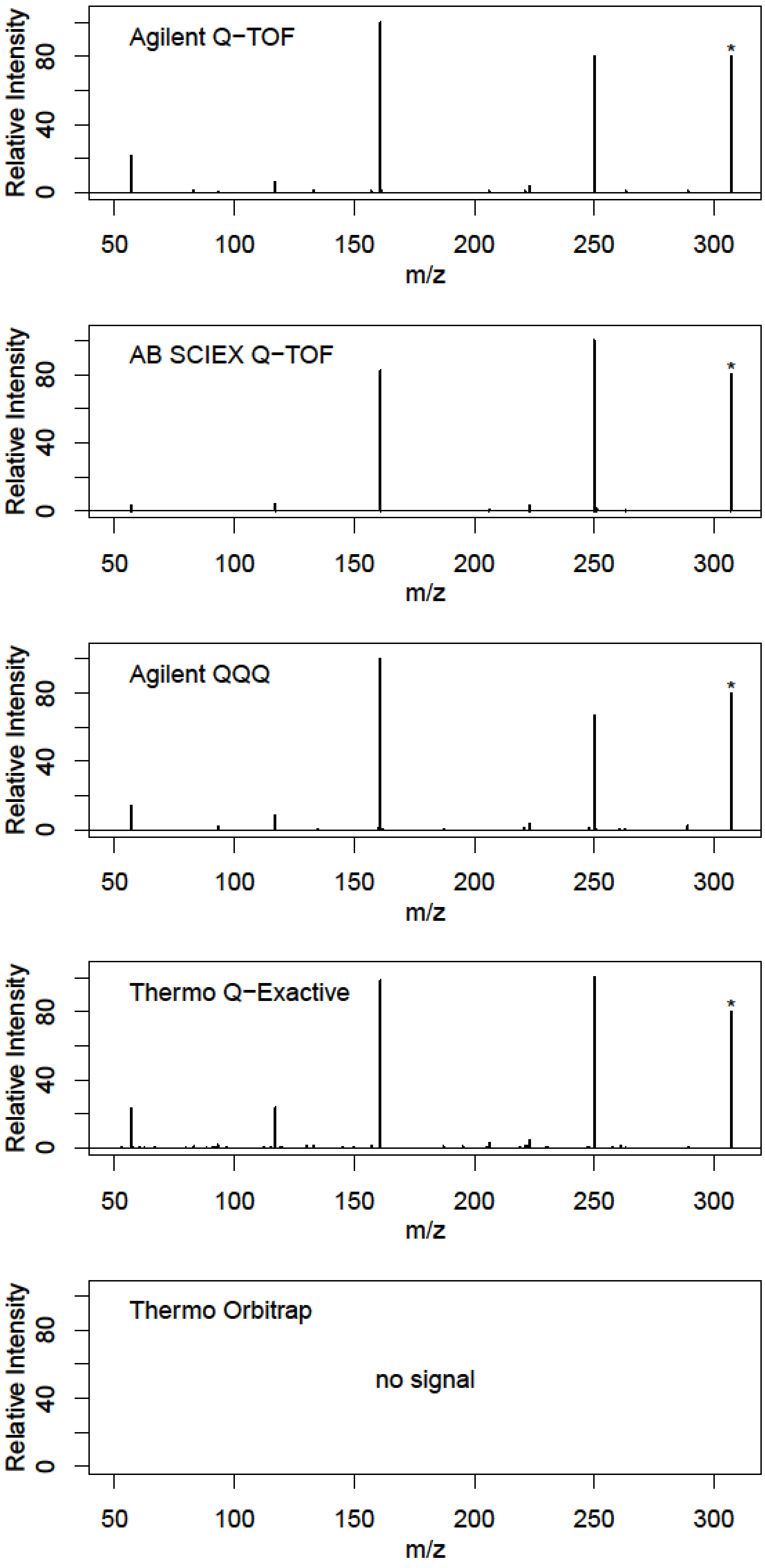
Supplementary Figure 27:

A comparison of MS^2 data for 2-Ethyl-2-Hydroxybutyric acid stored in METLIN to MS^2 data for 2-Ethyl-2-Hydroxybutyric acid generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in negative ionization mode.



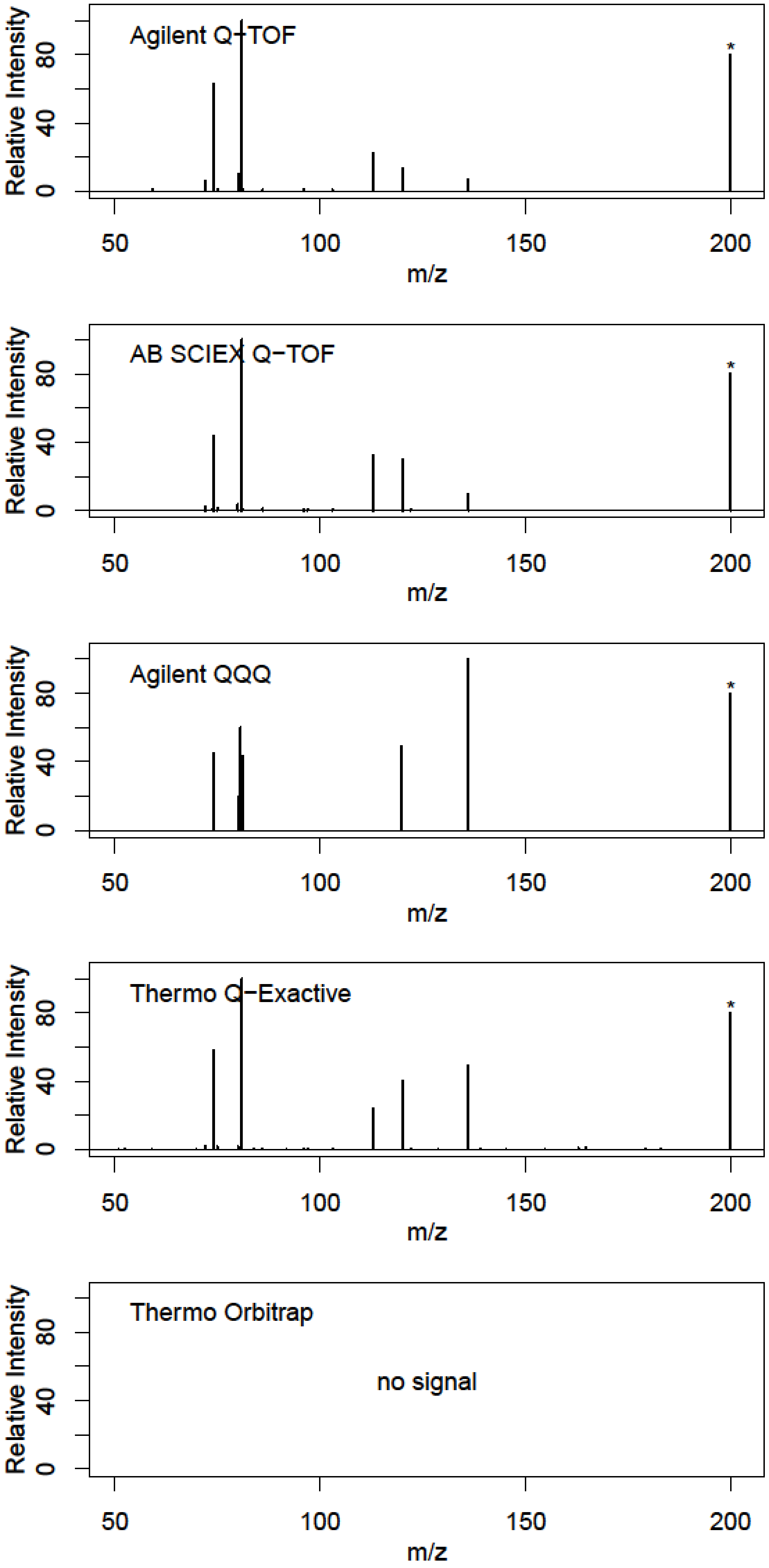
Supplementary Figure 28:

A comparison of MS^2 data for D-Ala-D-Ala stored in METLIN to MS^2 data for D-Ala-D-Ala generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in negative ionization mode.



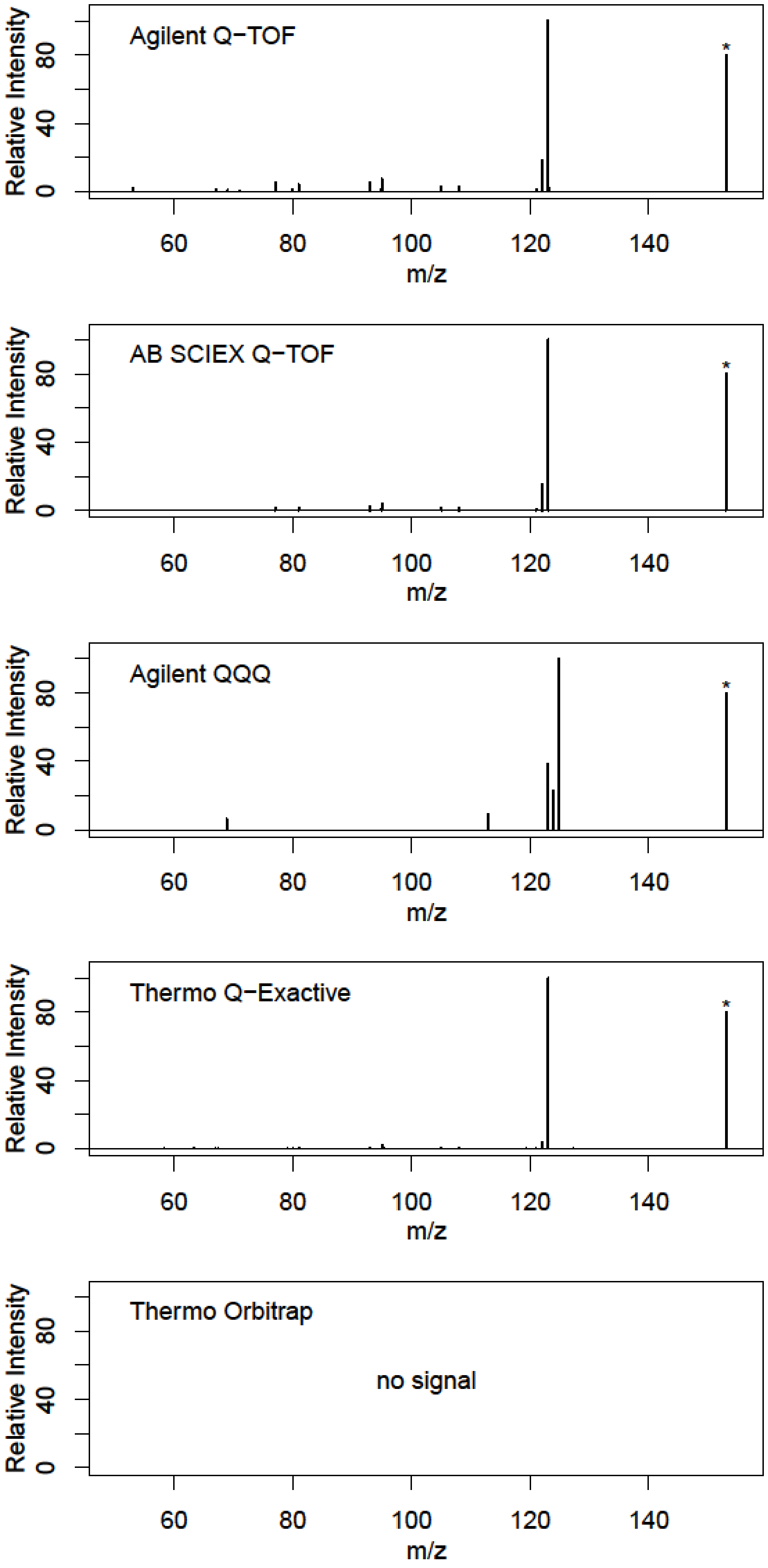
Supplementary Figure 29:

A comparison of MS^2 data for warfarin stored in METLIN to MS^2 data for warfarin generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in negative ionization mode.



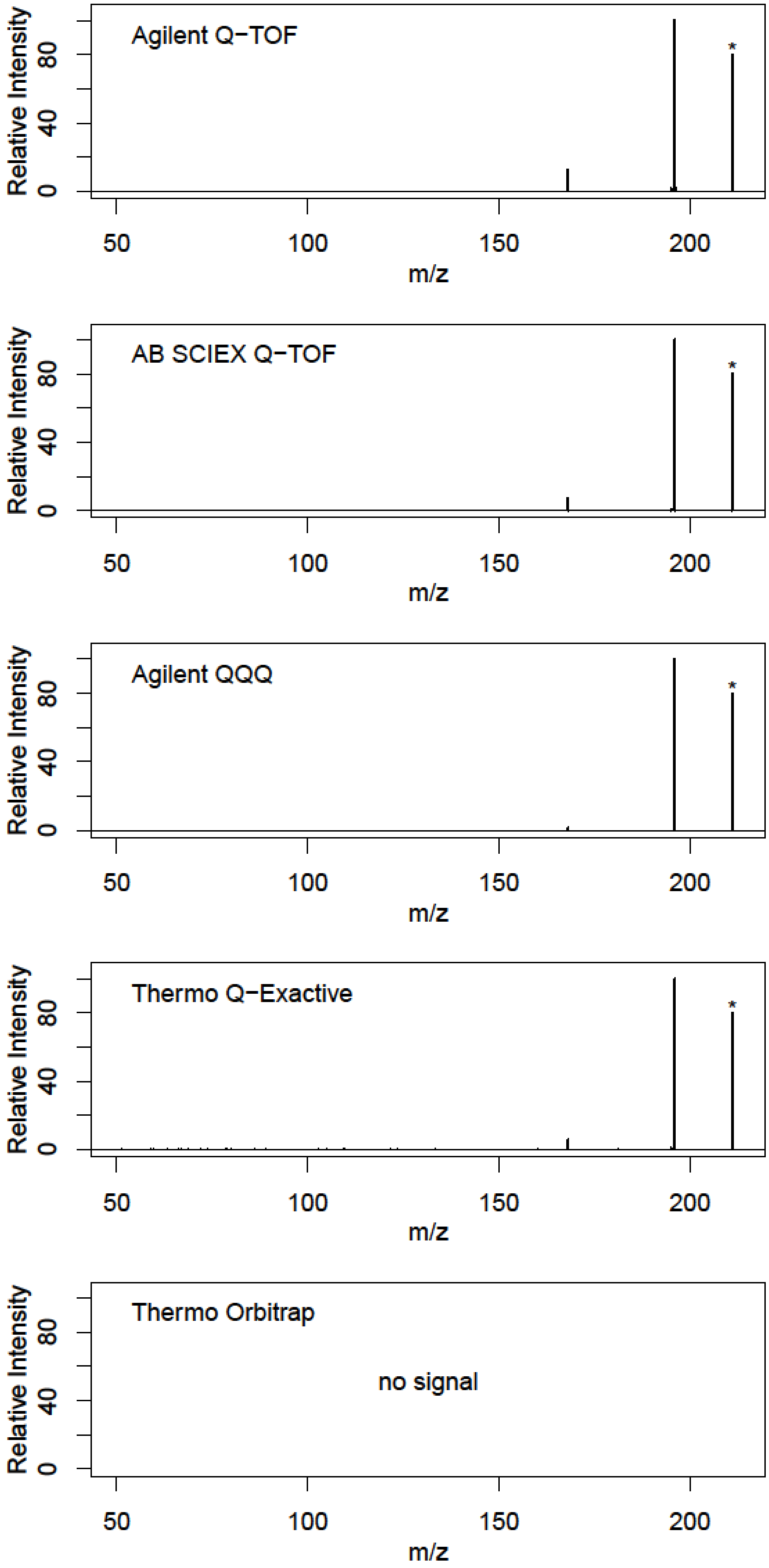
Supplementary Figure 30:

A comparison of MS^2 data for cysteine-S-sulfate stored in METLIN to MS^2 data for cysteine-S-sulfate generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in negative ionization mode.



Supplementary Figure 31:

A comparison of MS^2 data for 3,4-Dihydroxyphenyl ethanol stored in METLIN to MS^2 data for 3,4-Dihydroxyphenyl ethanol generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in negative ionization mode.



Supplementary Figure 32:

A comparison of MS^2 data for harmine stored in METLIN to MS^2 data for harmine generated from different instrument platforms. MS^2 data in METLIN were generated by using an Agilent 6510 Q-TOF with a CID of 20V. AB SCIEX Q-TOF data were generated from an AB SCIEX Triple TOF 5600 with a CID of 30V. Agilent QQQ data were generated from a 6460 Triple Quad with a CID of 10V. Q-Exactive data were generated from a Thermo Scientific Q-Exactive with an HCD of 50V. Orbitrap data were generated from a Thermo Scientific LTQ Orbitrap Velos with an HCD of 40V. All data were generated in negative ionization mode.

References

1. Yanes, O., *et al.* Metabolic oxidation regulates embryonic stem cell differentiation. *Nat Chem Biol* **6**, 411-417 (2010).
2. Yanes, O., Tautenhahn, R., Patti, G.J. & Siuzdak, G. Expanding coverage of the metabolome for global metabolite profiling. *Anal Chem* **83**, 2152-2161 (2011).
3. http://metlin.scripps.edu/spec_search.php